

RESEARCH ARTICLE

Selection and validation of reference genes for quantitative real-time PCR in the green microalgae *Tetraselmis chui*

Sonia Torres¹, Carmen Lama¹, Lalia Mantecón¹, Emmanouil Flemetakis², Carlos Infante^{1*}

1 Fitoplancton Marino, S.L., El Puerto de Santa María, Cádiz, Spain, **2** Laboratory of Molecular Biology, Department of Biotechnology, Agricultural University of Athens, Athens, Greece

* carlos.infante@easylgae.com



OPEN ACCESS

Citation: Torres S, Lama C, Mantecón L, Flemetakis E, Infante C (2021) Selection and validation of reference genes for quantitative real-time PCR in the green microalgae *Tetraselmis chui*. PLoS ONE 16(1): e0245495. <https://doi.org/10.1371/journal.pone.0245495>

Editor: Shu-Biao Wu, University of New England, AUSTRALIA

Received: June 5, 2020

Accepted: December 30, 2020

Published: January 14, 2021

Copyright: © 2021 Torres et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: This work was co-funded by the project A2F “Algae to Future” (number 267872) from the Research Council of Norway, and by the EU project Algae4a-b “Development of Microalgae-based novel high added-value products for the Cosmetic and Aquaculture industry” (H2020- MSCA-RISE-15; grant agreement 691102). Fitoplancton Marino, S.L. also provided support in the form of salaries

Abstract

Quantitative real-time reverse transcription PCR (RT-qPCR) is a highly sensitive technique that can be applied to analyze how genes are modulated by culture conditions, but identification of appropriate reference genes for normalization is a critical factor to be considered. For this reason, the expression stability of 18 candidate reference genes was evaluated for the green microalgae *Tetraselmis chui* using the widely employed algorithms geNorm, Norm-Finder, BestKeeper, the comparative ΔC_T method, and RefFinder. Microalgae samples were collected from large scale outdoor photobioreactors during the growing phase (OUT_GP), and during the semi-continuous phase at different times of the day (OUT_DC). Samples from standard indoor cultures under highly controlled conditions (IND) were also collected to complement the other data. Different rankings for the candidate reference genes were obtained depending on the culture conditions and the algorithm employed. After comparison of the achieved ranks with the different methods, the reference genes selected for samples from specific culture conditions were *ALD* and *EFL* in OUT_GP, *RPL32* and *UBCE* in OUT_DC, and *cdkA* and *UBCE* in IND. Moreover, the genes *EFL* and *cdkA* or *EFL* and *UBCE* appeared as appropriate combinations for pools generated from all samples (ALL). Examination in the OUT_DC cultures of genes encoding the large and small subunits of ADP-glucose pyrophosphorylase (*AGPL* and *AGPS*, respectively) confirmed the reliability of the identified reference genes, *RPL32* and *UBCE*. The present study represents a useful contribution for studies of gene expression in *T. chui*, and also represents the first step to set-up an RT-qPCR platform for quality control of *T. chui* biomass production in industrial facilities.

Introduction

Microalgae represent the most phylogenetically diverse group of organisms on Earth [1,2]. As photosynthetic organisms, they can utilize solar energy to convert inorganic CO₂ into a wide set of valuable biologically active compounds, including fatty acids, pigments, vitamins, and

for ST, CL, LM, and CI. The specific roles of these authors are articulated in the 'author contributions' section. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have read the journal's policy and have the following competing interests: ST, CL, LM, and CI are or were employees of Fitoplancton Marino, S.L. This does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

proteins and enzymes. Owing to the metabolic plasticity of microalgae, stress caused by changes in factors such as salinity, temperature, light and nutrients, can lead to the production and accumulation of unique metabolites [3,4]. Such molecules have neurological, antimicrobial, antifungal, antiprotozoal, antiviral, anti-inflammatory, antioxidant, anticancer and immune-stimulant activities, and this explains the high potential of microalgae for cosmetics, pharmaceuticals, feed and food, and as a source of and supplements for nutraceuticals [5–7].

The green algae genus *Tetraselmis* (Chlorodendrophyceae) comprises species that are among the most widely used in aquaculture feed for larval nutrition in molluscs, shrimp and fish, as well as for the enrichment of rotifers and *Artemia* nauplii [8,9]. More specifically, *Tetraselmis chui* (Butcher 1959) exhibits an appropriate content of proteins, lipids, carbohydrates and essential fatty acids and sterols, for the maintenance of aquaculture organisms [10,11]. In addition to its nutritional value, some studies suggest that *T. chui* supplementation can enhance the immune system and health in the organisms to which it is fed. For example, inclusion of *T. chui* in formulated feed enhances defense against oxidative stress and damages in Pacific white-leg shrimp (*Litopenaeus vannamei*) postlarvae [12], and improves immune-related parameters in the fish, gilthead seabream (*Sparus aurata*) [13,14]. Furthermore, the company Fitoplancton Marino, S.L. (El Puerto de Santa María, Cádiz, Spain) has been authorized since 2014 to market its lyophilized *T. chui* biomass as a novel food for human consumption in the European Union (EU) according to Regulation (EC) No 258/97. In 2017, commercialization by Fitoplancton Marino, S.L. of the lyophilized biomass, TetraSOD[®], for use as a food supplement for humans was also approved in the EU [15].

It has been stated that gene expression analysis by quantitative real-time reverse transcription polymerase chain reaction (RT-qPCR) represents an important tool for understanding biological and molecular mechanisms underlying cellular processes [16–18]. The comparison of the gene expression patterns across different experimental or biological conditions can provide valuable insights into the regulation of the metabolic pathways that help cells withstand or adapt to environmental change. In this context, RT-qPCR is one of the most commonly used approaches to measure gene expression levels owing to its accuracy, sensitivity, specificity, precision, real-time detection of the reaction progress, dynamic range, and the speed of analysis [19–22]. However, a range of factors can significantly diminish the accuracy and reliability of RT-qPCR data, such as quality and quantity of the mRNA templates, enzymatic efficiency of the reverse transcription and PCR amplification reaction, and differences between cells or tissues in the overall transcriptional activity and mRNA half-life [23,24]. To avoid the influence of these factors, they are usually controlled by normalizing gene expression to an internal control or reference gene. Such reference genes should have an expression unaffected by experimental factors, have minimal variation in expression between tissues and physiological states of the organism, and a threshold cycle similar to the target gene. However, studies have revealed that it is difficult to identify universal reference genes that do not vary under different experimental conditions or with cell type or organism, and hence it is essential to take this into account when identifying reference genes in order to ensure the accuracy of RT-qPCR analysis [24–26]. Several statistical algorithms have been developed to identify the most suitable reference genes under a given set of experimental or biological conditions in order to avoid artifacts due to normalization, and include geNorm [27], NormFinder [28], BestKeeper [29], the comparative ΔC_T method [30], and the web-based program RefFinder [31]. These tools have been carried out on a broad range of species, including whole organisms, specific tissues or cell types from animals [19–21,32,33], plants [34–37], and fungi [38,39]. Selection of reference genes has also been addressed in macroalgae [40–42] as well as in specific microalgae species, including the diatoms *Phaeodactylum tricornutum* [43], *Pseudo-nitzschia multistrata* and *P. arenysensis* [44], *Ditylum brightwellii* [45] and *Skeletonema marinoi* [46], the dinoflagellates

Prorocentrum minimum [47], *Symbiodinium* [48], *Alexandrium catanella* [49] and *Karenia mikimotoi* [50], the haptophyte *Isochrysis zhangjiangensis* [51], the eustigmatophyte *Nannochloropsis* [52,53], the raphidophyte *Heterosigma akashiwo* [54], or the green algae *Volvox carteri* [55], *Closterium ehrenbergii*, [56], *Chlamydomonas* [17,57], and *Tetraselmis suecica* [58]. However, to our knowledge, there has been no study validating reference genes for the green microalgae species *T. chui*.

In the present study, the suitability of a set of 18 different candidate reference genes for normalization in RT-qPCR assays in the microalgae species *T. chui* were evaluated. The candidate genes were chosen based on their previous assessment in other organisms, particularly in other microalgae. The candidate genes included, 18S rRNA (*18S*), actin (*ACT*), aldolase A (*ALD*), alpha tubulin (*aTUB-1* and *aTUB-2*), beta tubulin (*bTUB*), cyclin dependent kinase A (*cdkA*), elongation factor-like (*EFL*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), histone 2A (*His2A*), beta-ketoacyl-ACP synthase (*KAS*), large subunit of the ribulose-1,5-biphosphate carboxylase (*rbcL*), ubiquitin conjugating enzyme (*UBCE*), phosphoglycerate kinase (*PGK*), ribosomal protein S10 (*RPS10*) and L32 (*RPL32*), and eukaryotic translation initiation factor 2 (*eIF2-1* and *eIF2-2*). *T. chui* microalgae samples were first collected from outdoor photobioreactors dedicated to the industrial production of biomass operated in semi-continuous mode. Sampling was performed from initial inoculation in photobioreactors until the cultures reached the appropriate cell density for daily harvesting and spanned 18 days. Thereafter, samples were drawn from outdoor photobioreactors at five different hours of the diurnal cycle from dawn to dusk. The collected samples represented microalgae exposed to strong differences in light irradiance (ranging from 3 to 1300 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ of photosynthetically active radiation (PAR)) and also in culture temperature (ranging from 10.6 to 22.8°C). Further samples were obtained from indoor cultures grown under stable temperature and light irradiance and spanned the growth curve from the initial inoculation to the stationary phase. Thereafter, the transcript abundance of the candidate reference genes were determined by RT-qPCR, and analyzed using geNorm, NormFinder, BestKeeper, the comparative ΔC_T method, and RefFinder to identify reference genes suitable for gene expression normalization. Finally, the expression of the genes encoding for the large and small subunits of ADP-glucose pyrophosphorylase (*AGPL* and *AGPS*, respectively) were examined by using the most stable reference genes selected.

Materials and methods

Strain and culture conditions

T. chui strain CCFM-03 (belonging to the Culture Collection of Fitoplankton Marino, S.L.) is cultured for industrial production of biomass in the facilities of the company Fitoplankton Marino, S.L. For the purpose of this study, 4000 L closed tubular outdoor photobioreactors were sampled. Microalgae cells were grown in seawater with f/2 culture medium [59] throughout the production cycle, including the first indoor stages during culture up-scaling through to outdoor cultures. The composition of f/2 medium was: 8.82×10^{-4} M NaNO_3 , 3.62×10^{-5} M $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$, 1.06×10^{-4} M Na_2CO_3 , 1.17×10^{-5} M $\text{FeCl}_3 \cdot 6(\text{H}_2\text{O})$, 1.17×10^{-5} M $\text{Na}_2(\text{EDTA}) \cdot 2(\text{H}_2\text{O})$, 3.93×10^{-8} M $\text{CuSO}_4 \cdot 5(\text{H}_2\text{O})$, 2.6×10^{-8} M $\text{Na}_2\text{MoO}_4 \cdot 2(\text{H}_2\text{O})$, 7.65×10^{-8} M $\text{ZnSO}_4 \cdot 7(\text{H}_2\text{O})$, 4.2×10^{-8} M $\text{CoCl}_2 \cdot 6(\text{H}_2\text{O})$, 9.1×10^{-7} M $\text{MnCl}_2 \cdot 4(\text{H}_2\text{O})$, 2.96×10^{-7} M vitamin B1, 2.05×10^{-9} M vitamin H, and 3.69×10^{-10} M vitamin B12. In outdoor cultures, the microalgae were exposed daily to ambient conditions, with temperature and light intensity being probably the most significant parameters that changed (S1 Table). Culture conditions that were strictly controlled included pH (maintained at 7.5–8.5 by on-demand CO_2 injections) and nitrate content, which was maintained over 500 μM . Indoor cultures of microalgae

were maintained under controlled conditions at 22°C in a thermoregulated room with a photoperiod of 24 h of cool fluorescent white light, and ~150 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ of PAR. All microalgae cultures were continuously aerated using compressed air containing 2% CO₂.

Microalgae samples

Three different conditions were tested in the study: i) highly controlled indoor cultures (IND), ii) outdoor cultures during the growth phase (OUT_GP), that was, from initial inoculation of the photobioreactors until the cultures reached an appropriate cell density for semi-continuous cultivation, and iii) outdoor cultures during the semi-continuous phase (OUT_DC), in which the culture was maintained in exponential growth over ~2 months by daily harvesting of biomass followed by addition of nutrients to restore adequate levels of nitrate, salts and vitamins.

For the IND cultures, three 5 L flasks were inoculated on day 1 (D1) at an average initial cell density of 85×10^3 cells/ml (1/20 dilution of a well grown source culture), which was determined using a haematocytometer (Neubauer counting chamber) and a microscope. Microalgae samples were withdrawn from each of the triplicate flasks for cell density determination and RNA isolation on D1, D4, D7, D10 and D13. Samples from D4 and D7 corresponded to the exponential growth phase and samples from D10 and D13 corresponded to the stationary growth phase (S1 Fig). A total of 15 samples were collected (three per time point) for gene expression analysis. Collection of samples for both cell density determination and RNA isolation occurred at 9:00 a.m. on the five sampling days.

OUT_GP samples were collected from three 4000 L photobioreactors of the same production unit after inoculation on D1 (with an average cell density of 20×10^3 cells/ml), and then at D4, D7, D11, and finally at D18 (the day before the commencement of the semi-continuous cultivation mode) during the exponential growth phase. In all instances, the samples were taken at 1:00 p.m. Samples were collected for cell density determination (S2 Fig) and 15 samples (three per time point) were collected for RNA isolation and further gene expression analysis.

OUT_DC samples were collected from three 4000 L photobioreactors of the same production unit and during the semi-continuous cultivation mode. Samples were withdrawn 13 days after the commencement of the semi-continuous phase (approximately 30 days after inoculation of the photobioreactors) for RNA isolation at five different times of the diurnal cycle: 7:00 a.m., 10:00 a.m., 1:00 p.m., 4:00 p.m. and 7:00 p.m. Thus, a total of 15 samples (three per time point) were processed for gene expression analysis. The average cell density at 7:00 a.m. (the first sample of a given was 7.4×10^6 cells/ml).

The temperature of cultures in OUT_GP and OUT_DC was automatically controlled with a probe at each sampling time, and PAR was also recorded using a radiometer sensor LI-250A Light Meter (LICOR Biosciences). The data recorded is presented in S1 Table.

Total RNA isolation and cDNA synthesis

In each sample, an appropriate volume of culture was drawn to give ~ 4×10^6 cells for RNA isolation. Cells were centrifuged at 5,000 g for 5 min, and the supernatant was removed. Microalgae cells were homogenized using stainless steel beads (0.2 mm, Next Advance) for 3 min at speed 10 in a Bullet Blender[®] 24 (Next Advance). Total RNA was extracted using a NucleoSpin[®] Plant II kit (Macherey-Nagel) following the manufacturer's instructions. The extracted RNA was treated twice with DNase I to avoid further PCR amplification of residual traces of genomic DNA. RNA samples were quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and the quality was checked in agarose gels.

Total RNA (1 μ g) from each sample was reverse-transcribed using an iScript™ cDNA Synthesis kit (Bio-Rad) in a reaction volume of 20 μ L according to the manufacturer's protocol. All cDNA reactions were finally diluted 10-fold by adding 180 μ L of nuclease-free water. The absence of genomic DNA contamination was confirmed by direct PCR amplification of two randomly selected RNA samples in the absence of cDNA synthesis.

Selection of candidate reference genes for RT-qPCR

An initial set of 16 genes involved in different cellular processes were pre-selected to assess their suitability as reference genes for RT-qPCR analysis. The selection of candidate reference genes was based on previous gene expression studies of other microalgae species and other organisms. The nucleotide sequence corresponding to *18S* and *rbcl* from *T. chui* was directly retrieved from GenBank/EMBL/DDBJ database (S2 Table). The nucleotide sequences of the other candidate genes were identified in the available transcript sequences (~22,600) for the *T. chui* strain PLY429 in the iMicrobe database (www.imicrobe.us) after annotation with the AutoFACT tool [60]. EditSeq v8.1.3 (DNASTAR) was used to identify the coding sequences of candidate genes as well as the deduced protein sequences, and then gene identities were confirmed with BLASTp (S2 Table). Two different coding sequences were found in *T. chui* for both alpha tubulin (*aTUB-1* and *aTUB-2*) and eukaryotic translation initiation factor 2 (*eIF2-1* and *eIF2-2*), and all four genes were included in this study. The isolation of the target genes *AGPL* and *AGPS* followed the same general strategy outlined above.

Primer design and RT-qPCR

Specific primer pairs (Table 1) for the 18 selected candidate reference genes (and also for the target genes *AGPL* and *AGPS*) were designed using Oligo v7.60 software (Molecular Biology Insights). To ensure maximum specificity and efficiency in PCR amplification, highly stringent conditions were initially selected for primer design. For each primer pair these were: ΔG of the most stable 3'-dimer > -3.5 Kcal/mol, ΔG of the most stable dimer overall > -7.5 Kcal/mol, no hairpins with $\Delta G < -1$ Kcal/mol and a melting temperature (T_m) $> 40^\circ\text{C}$, primer efficiency (PE) > 480 , PE of false priming sites < 100 , optimal annealing temperature (T_a) of the PCR product between 62 – 65°C , primer lengths of 20–28 nucleotides, and amplicon length ranging between 100–200 base pairs. When most of the criteria were successfully met, the recommended temperature for annealing/extension in PCR was 68°C . All designed primers were finally ordered from a commercial supplier (Eurofins). Appropriate performance of each primer pair was initially tested by PCR amplification of the target amplicons employing the same conditions described below for RT-qPCR. Specificity of each primer pair was verified by melting curve analysis from 70°C to 95°C with a ramp speed of 0.5°C every 10 s. Single, sharp peaks were obtained in all instances, thus ruling out amplification of non-specific products or primer dimer artifacts (S3 Fig). To determine real-time PCR efficiency (E) of each primer pair, standard curves were prepared from serial dilutions of cDNA (from 100 ng to 0.01 ng) and the results plotted in Excel against their C_T values (S4 Fig). Then, slopes of the linear regressions were determined applying the equation $E = 10^{-1/\text{slope}}$ as previously reported [61]. The linear correlation coefficient (R^2) of each primer pair was also determined in Excel. The PCR products were analyzed by standard agarose gel electrophoresis (2.5% w/v in TAE 1X) and clear DNA bands of the expected sizes were obtained (S5 Fig).

RT-qPCR was conducted using a CFX96™ Real-Time PCR Detection System (Bio-Rad). Each 10 μ L reaction contained 5 μ L of 2X iQ™ SYBR® Green Supermix (Bio-Rad), 300 nM of both forward and reverse primers (0.3 μ L of a 10 μ M stock each), 2 μ L of cDNA (corresponding to the cDNA retrotranscribed from 10 ng of RNA), and 2.4 μ L of nuclease-free water. All

Table 1. List of primers used for RT-qPCR.

CANDIDATE REFERENCE GENES					
Gene symbol	Primer sequence	Amplicon length (bp)	Optimal T _a (°C) ^a	PCR efficiency (E)	Correlation coefficient (R ²)
<i>ACT</i>	F: 5'-AGAAGACCTATGAGCTGCCCGACG-3'	168	64.0	1.06	1.00
	R: 5'-GGTCCTTACGGATATCGACATCGCACT-3'				
<i>EFL</i>	F: 5'-CCGGCAGATCAAGGTCGGCTAC-3'	171	63.9	1.05	1.00
	R: 5'-AGGGCTTGAAGACAATGGTGGATACCTC-3'				
<i>His2A</i>	F: 5'-TCATCGTTTGTGTAAGAACCGTGTGAC-3'	146	62.3	1.06	1.00
	R: 5'-TGCCTTACCTTCAGATCCTTAGACG-3'				
<i>ALD</i>	F: 5'-TGTACAAGCCCGCAACGTGA-3'	176	64.9	1.03	0.99
	R: 5'-CTTGATGACGCCGTAGCCGAT-3'				
<i>cdkA</i>	F: 5'-ACCGCAGAACTACTGATTGACCGT-3'	123	61.2	1.07	1.00
	R: 5'-CGGTACCACAGAGTCACAACCTCGT-3'				
<i>GAPDH</i>	F: 5'-GCCATCGCGCTCATCTACCCCGAA-3'	110	62.4	1.07	0.99
	R: 5'-TTCGGTTCACCTCAAAGACACAATCGGT-3'				
<i>KAS</i>	F: 5'-CGATGCACACCACATGACCGACC-3'	172	65.5	0.95	1.00
	R: 5'-CGCCTGCTTCATCGCCTTGACC-3'				
<i>rbcl</i>	F: 5'-ACGTAATTTCACAAGCTTTCATGCGTTGG-3'	140	57.0	1.05	1.00
	R: 5'-CATCATTCTTCACAAGTCCCAGCCGTT-3'				
<i>RPL32</i>	F: 5'-CTCCAACAAGAAGACCCGCCACC-3'	131	62.3	1.02	1.00
	R: 5'-TGGACACGTTCTTGGGACCTC-3'				
<i>aTUB-1</i>	F: 5'-ACTCGACTGCTATCGCTGAAGTCTTCTCA-3'	118	62.0	0.99	1.00
	R: 5'-GAGAAGTCCGCTCCTCCATACCCTCA-3'				
<i>aTUB-2</i>	F: 5'-CCCTCGCGTGTGTTTCATCGAC-3'	109	62.9	1.01	0.99
	R: 5'-GTCTTCTTCCGGAGATCAGCTGCTCA-3'				
<i>bTUB</i>	F: 5'-GCCAGATCTCCGCCCTGACAAC-3'	133	63.0	1.02	1.00
	R: 5'-GACTCGGCCTCCTTACGGACCAC-3'				
<i>18S</i>	F: 5'-GGGGGAGTATGGTCGCAAGGCTGAA-3'	183	62.4	0.96	1.00
	R: 5'-AACTAAGAACGGCCATGCACCACCAC-3'				
<i>UBCE</i>	F: 5'-CCAAACATCAACAGCAACGGCAGCA-3'	149	64.1	0.98	1.00
	R: 5'-TGCGCAATCTCGGGCACCAG-3'				
<i>PGK</i>	F: 5'-CAACAAGTGCAGACAAGATCATCATTGGC-3'	178	63.7	1.01	1.00
	R: 5'-AACGACGACATCGGTAGGGAGCAA-3'				
<i>RPS10</i>	F: 5'-CAAGAAGAACCGCCGCGAGGTGT-3'	172	63.0	1.02	1.00
	R: 5'-CCACGCAAAACGCTCAGTGACCAG-3'				
<i>eIF2-1</i>	F: 5'-CCTTTGGAAAGATTCGAGCAATGACGGAC-3'	129	65.1	1.04	1.00
	R: 5'-CCGCCACCGTGAACCTCGTCT-3'				
<i>eIF2-2</i>	F: 5'-CTTTGATCGTGAACCGTGGTGTGCT-3'	154	63.2	1.00	0.99
	R: 5'-ACACCCTTGATTCCAGCCACTACGAC-3'				
TARGET GENES					
Gene symbol	Primer sequence	Amplicon length (bp)	Optimal T _a (°C)	PCR efficiency (E)	Correlation coefficient (R ²)
<i>AGPL</i>	F: 5'-CCGCCACCATCACTCCCGAAT-3'	164	63.2	1.06	1.00
	R: 5'-CCGCAGAGGTCTGATAGTCCAT-3'				
<i>AGPS</i>	F: 5'-GACTTCTCATCCTCTCCGGCGACCA-3'	163	66.5	0.96	1.00
	R: 5'-TCCGCCCGCTGTCGTGATC-3'				

^a T_a represents annealing temperature.

<https://doi.org/10.1371/journal.pone.0245495.t001>

reactions were run in duplicate, and the mean threshold cycle (C_T) of each sample (S3 Table) was used for further calculations. Relative transcript levels of the target genes were determined using the $2^{-\Delta\Delta C_T}$ method [62]. The thermal cycling profile included an initial incubation at 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and 68°C for 30 s.

Analysis of expression stability of the candidate reference genes

Five different approaches were employed to determine the expression stability of the candidate reference genes. The geNorm, or pairwise comparison approach, ranks candidate genes according to their expression stability [27]. The gene stability measure, or M , is determined for a given reference gene as the average pairwise variation for that gene in relation to the remaining tested reference genes. The gene with the highest M is stepwise excluded, with a new M value being repeatedly calculated for the remaining genes until a final pair of genes with the lowest M (the most stable) are obtained. Moreover, this Excel-based software determines the optimal number of reference genes required for an accurate normalization of real-time PCR data. For this, the pairwise variation (V_n/V_{n+1}) was analyzed between the normalization factors NF_n and NF_{n+1} ; V_n values were calculated by stepwise inclusion of additional reference genes until the (n+1) gene did not contribute significantly to the newly determined normalization factor. NormFinder software (also Excel-based) is a model-based approach that ranks the candidate reference genes according to their minimal combined inter- and intra-group expression variation [28]. Such variations are combined to give a stability value, with the lowest value corresponding to the most stable expression. BestKeeper (an Excel-based statistical method) calculates the standard deviation (SD) and coefficient of variation (CV) from the input raw C_T values, with genes exhibiting the lowest SD and CV being considered as the most stable [29]. Reference genes with SD higher than 1 are considered as inconsistent and are excluded. The comparative ΔC_T algorithm uses the average SD values to rank the stability of all candidate reference genes, and the reference gene with the lowest SD is considered as the most stable [30]. Finally, as a complementary analysis to assess reference gene stability, the web-based tool RefFinder was employed to integrate the results from the four previous software packages, generating a comprehensive rank list of candidate genes [31].

Statistical analysis of RT-qPCR data

Statistical analyses to determine relative gene expression variations in OUT_DC, both for *AGPL* and *AGPS*, were conducted using Prism 6 (GraphPad Software). Changes in transcript abundance were obtained after: i) normalization with the two most stable reference genes using the combined outputs of geNorm, NormFinder, BestKeeper, and the comparative ΔC_T method for the OUT_DC condition (*RPL32* and *UBCE*), ii) normalization with the two genes exhibiting the highest stability with RefFinder in OUT_DC (*RPS10* and *rbcl*), or iii) normalization with the two genes exhibiting the lowest stability with all the employed methods in OUT_DC (*PGK* and *eIF2-1*). The relative gene expression levels were also determined for *AGPL* and *AGPS* after normalization with the most stable genes found in the global analysis of all samples and conditions (ALL) based on the combined results of geNorm, NormFinder, BestKeeper, and the comparative ΔC_T method (*EFL* and *cdkA*), as well as with the combination of *EFL* (top-ranked by RefFinder) with another of the most stably-expressed genes found by RefFinder (*ACT*). In all instances, data were analyzed using the Friedman test (non-parametric one-way ANOVA), and when significant, the Dunn's multiple comparison test was performed. Significance was accepted for $P < 0.05$.

Results

Expression levels of the candidate reference genes

The expression stability of the candidate reference genes in *T. chui* was first assessed through the analysis of C_T values in real-time PCR (a high C_T means that the gene has a low expression level, and vice versa). Variation of data was compared between OUT_GP, OUT_DC and IND conditions, and with all data (ALL) being analyzed together (Fig 1). All the examined genes exhibited amplification prior to cycle n° 40 of the PCR amplification profile, with *eIF2-1* always showing the highest C_T (in some instances close to 33) across all conditions. In contrast, the lowest C_T values (and hence the highest expression levels) were observed for *18S*, and in some samples it was slightly higher than 9. The smallest variation of C_T values in OUT_GP corresponded to *EFL* and it ranged from 15.73 to 17.58, a difference of 1.85. In OUT_DC and IND, *RPL32* (from 17.37 to 18.27, and a difference of 0.90) and *ALD* (from 20.48 to 21.54, and a difference of 1.06), respectively had the smallest variation. In the global analysis ALL, the lowest range of C_T values was 16.90 to 19.13 and corresponded to *RPL32*. In contrast, the highest range of C_T values in each of the three conditions tested, and also in ALL, was observed for

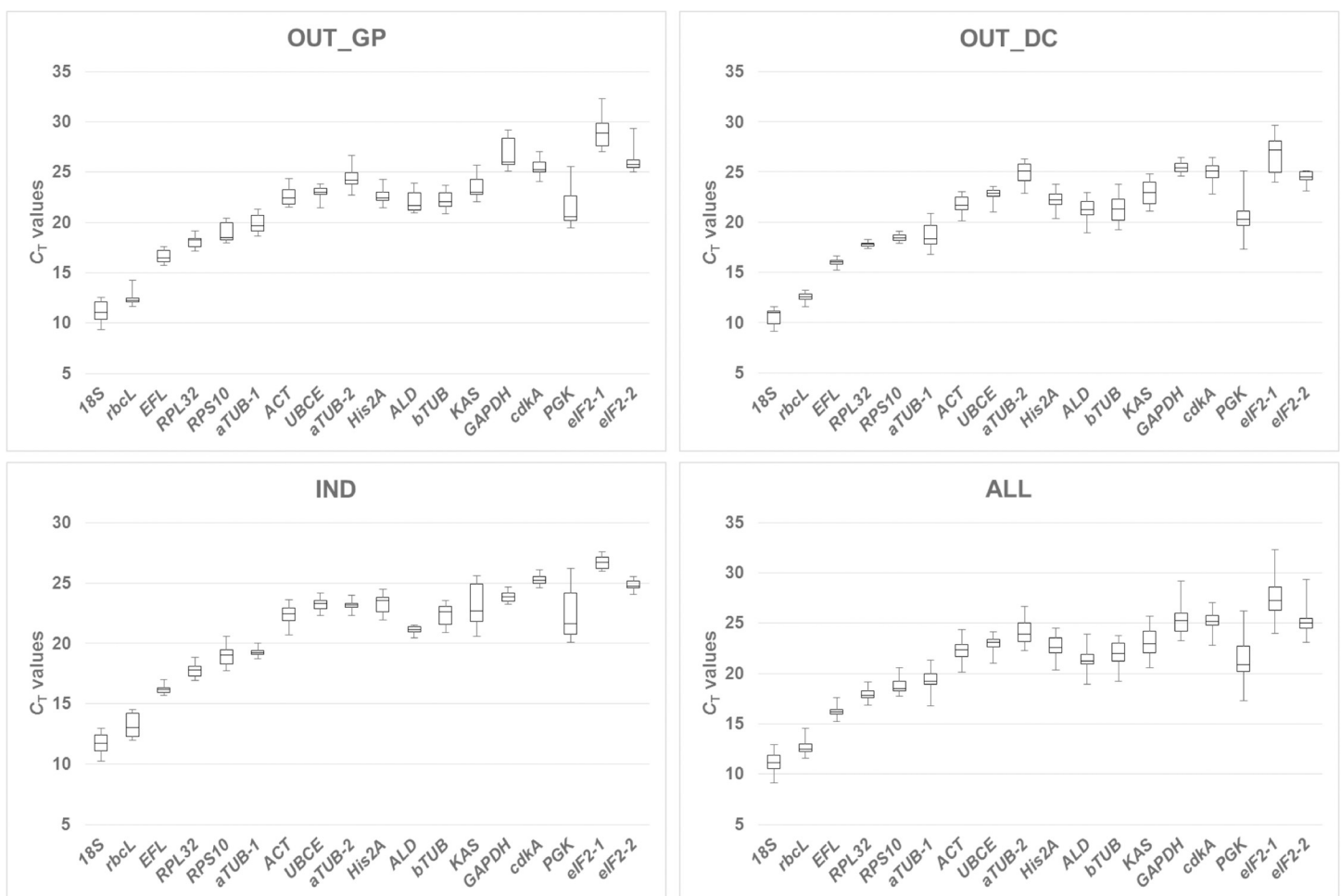


Fig 1. C_T values of candidate reference genes in all tested microalgae samples and conditions. The thick central line in boxes represents the median value, whereas the lower outline and the upper outline in boxes show the quartile 1 (Q1) and 3 (Q3), respectively. Whiskers are set at 1.5 times the interquartile range (IQR) both above Q3 and below Q1.

<https://doi.org/10.1371/journal.pone.0245495.g001>

PGK; 6.06 for OUT_GP, 7.80 for OUT_DC, 6.12 for IND and 8.92 for ALL. High variations in C_T values across samples and conditions indicate unsuitable reference genes. In the present study all 18 candidate reference genes were analyzed further.

Expression stability of reference genes

geNorm analysis. As a starting point, the average expression stability (M) values of the 18 candidate reference genes were determined using the geNorm algorithm (Fig 2). Lower M values indicate higher stability. In all instances, candidate genes included in this study displayed an M value below the geNorm threshold of 1.5, thus revealing all of them as suitable for gene expression normalization. The genes with the most stable expression differed with condition and *ALD* and *KAS*, *RPL32* and *RPS10*, and *cdkA* and *UBCE* displayed the lowest M values in OUT_GP, OUT_DC and IND samples, respectively. In ALL, the combination of the genes *EFL* and *RPS10* was the best. As a whole, the least stable genes were *eIF2-1* and *PGK*, although unexpectedly the former had the most stable expression in IND samples. To determine the optimal number of reference genes that would be required for an accurate normalization of

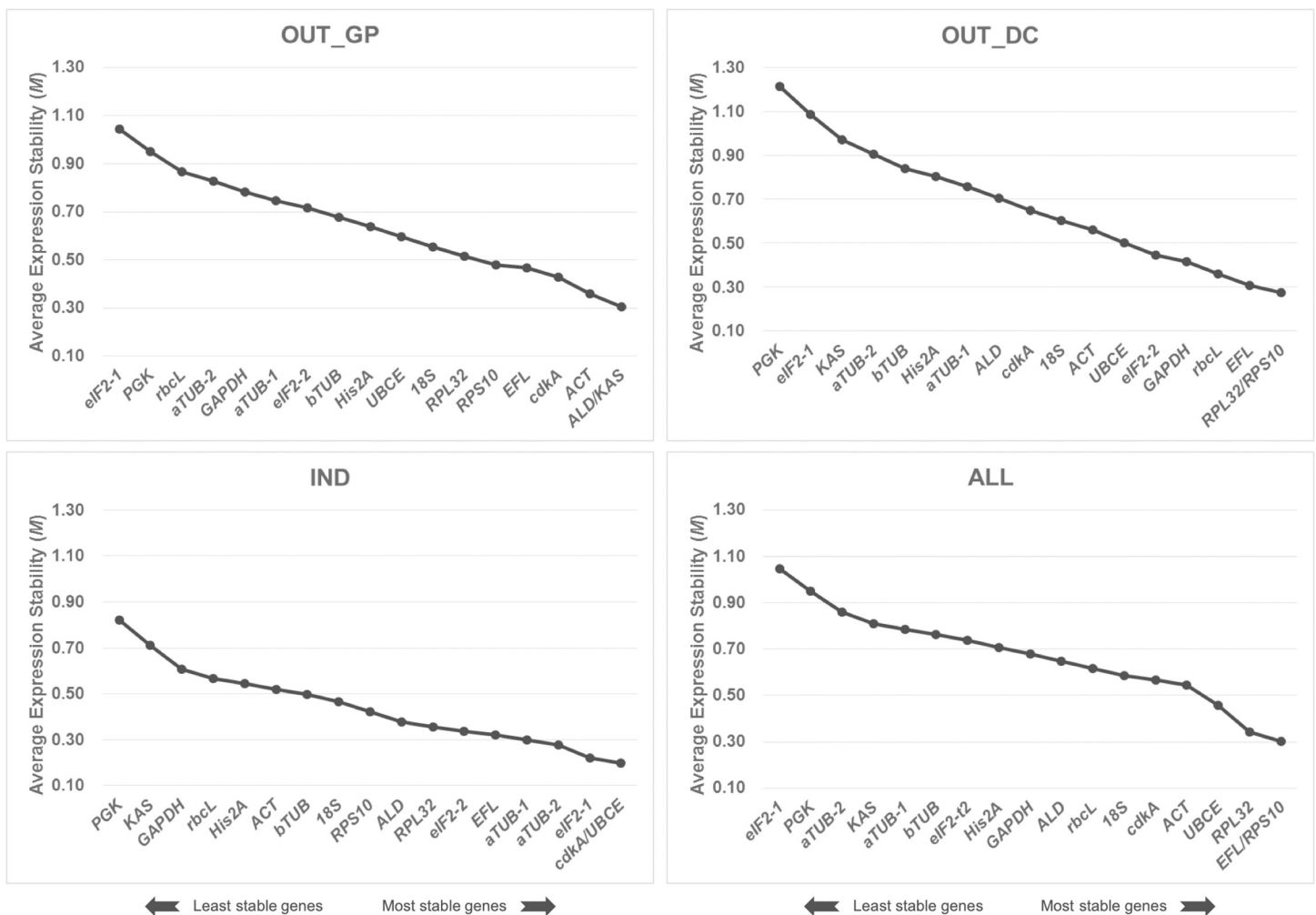


Fig 2. Global ranking of the 18 candidate reference genes using geNorm analysis. Average expression stability (M) was calculated by stepwise exclusion of the least stable gene identified by geNorm software in all tested samples (a total of 45) and conditions (OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: Indoor cultures, ALL: Combination of the three previous conditions). A lower M value indicates a more stable expression pattern.

<https://doi.org/10.1371/journal.pone.0245495.g002>

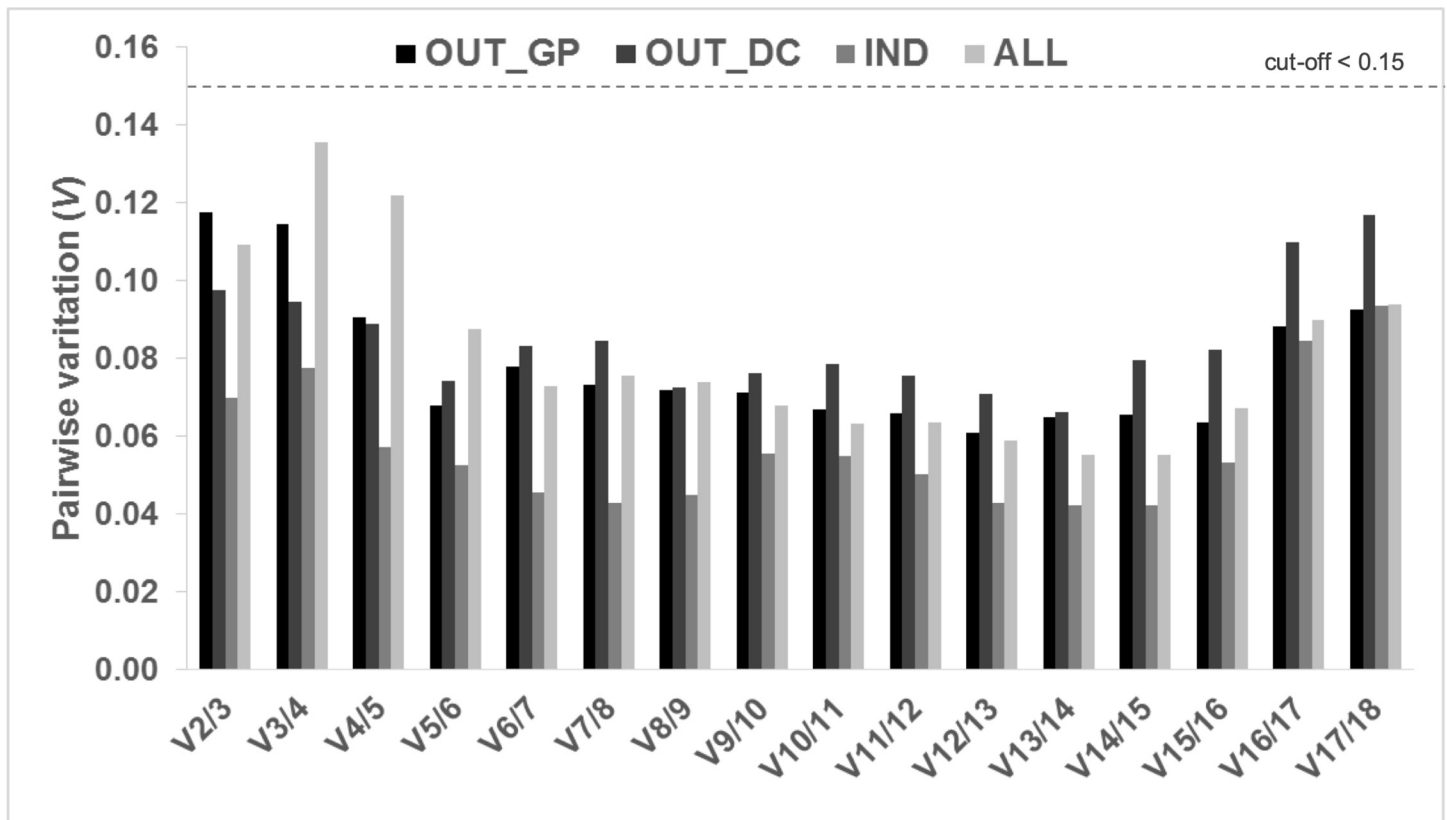


Fig 3. Determination of the optimal number of reference genes for accurate normalization of gene expression analysis. Pairwise variation was calculated in all tested samples (a total of 45) and conditions (OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: indoor cultures, ALL: Combination of the three previous conditions). The pairwise variation (V_n/V_{n+1}) was analyzed between the normalization factors NF_n and NF_{n+1} .

<https://doi.org/10.1371/journal.pone.0245495.g003>

gene expression analysis, pairwise variation values (V) were calculated with the cut-off of suitability set at 0.15, and below this value the addition of an additional internal reference gene does significantly improve normalization. As shown in Fig 3, the V2/3 was below the cut-off for all the experimental conditions studied and for the global analysis ALL. Thus, at least two genes were predicted as necessary for optimal normalization of gene expression levels in our experimental set-up.

NormFinder analysis. The stability of the 18 candidate reference genes evaluated with NormFinder software (Fig 4) should prevent misinterpretations of results owing to artificial selection of co-regulated genes. The stability ranking of candidate reference genes was significantly different between conditions. The top three stable candidate genes were *EFL*, *cdkA*, and *RPS10* in OUT_GP, while *UBCE*, *eIF2-2*, and *GAPDH* were ranked top in OUT_DC. In IND, *eIF2-1* was the most stable gene followed by *UBCE* and *RPS10*, although for OUT_GP, OUT_DC and ALL (pool of samples from all conditions) it was the least stable gene together with *PGK*. Finally, *cdkA* was the most stable gene in ALL followed by *RPS10* and *ACT*.

BestKeeper analysis. The BestKeeper algorithm is a Microsoft Excel-based software and calculates the standard deviation (SD) and the coefficient of variation (CV) based on C_T values. Candidate reference genes with lower SD and CV have higher stability and are more suitable for normalization and genes with $SD > 1$ are considered unstable and unsuitable for normalization. The results of the BestKeeper analysis are shown in Table 2. The top-ranked candidate genes in OUT_GP exhibited significantly higher SD (~2-fold) than the top-ranked genes in

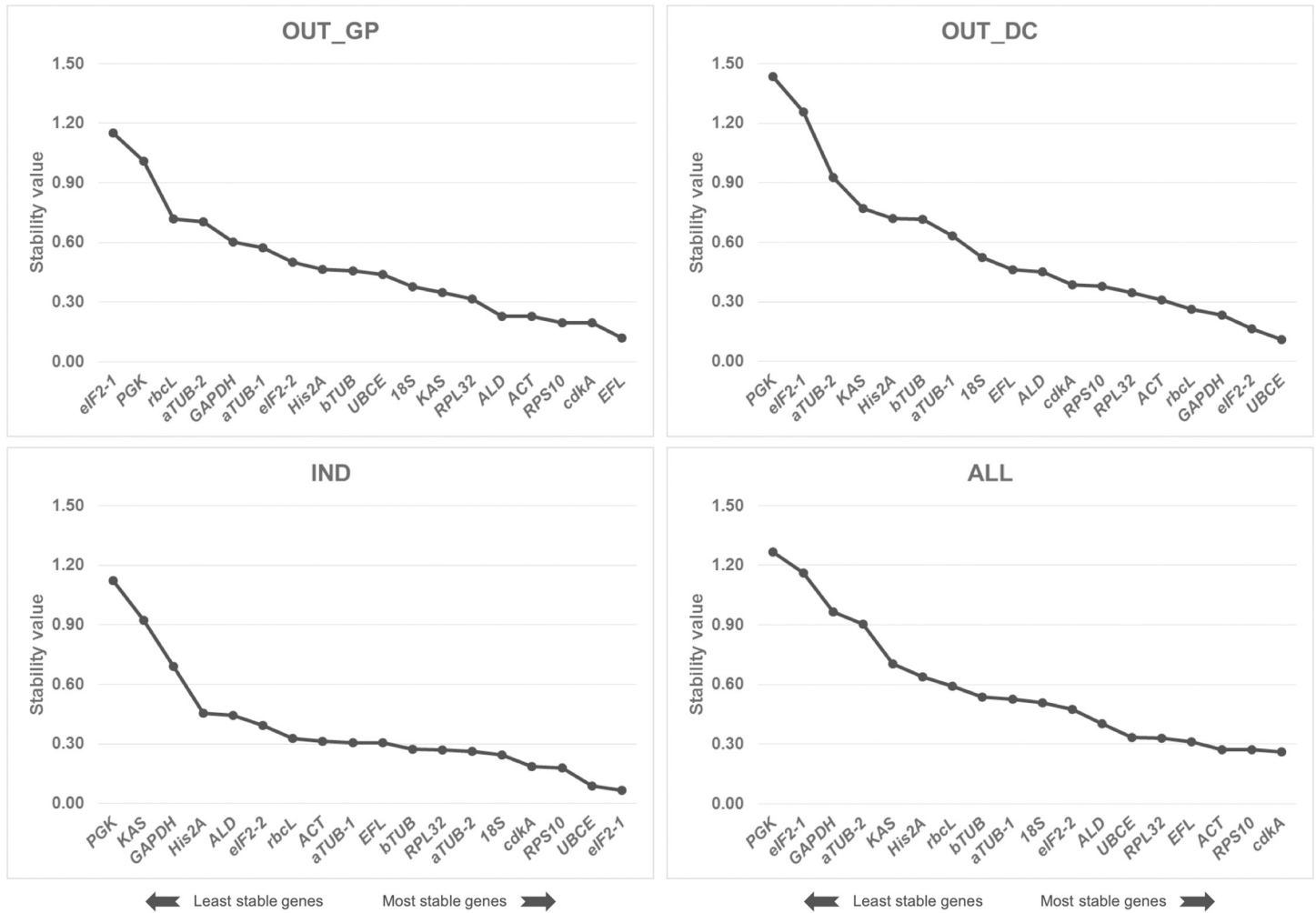


Fig 4. Global ranking of the 18 candidate reference genes using NormFinder analysis. The expression stability value was determined in all tested samples (a total of 45) and conditions (OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: Indoor cultures, ALL: Combination of the three previous conditions). A lower value indicates a more stable expression pattern.

<https://doi.org/10.1371/journal.pone.0245495.g004>

OUT_DC and IND. Based on SD values, the most stable genes were *UBCE* in OUT_GP, *RPL32* in OUT_DC, and *EFL* in IND and ALL. Moreover, *EFL* was also ranked among the top three genes in OUT_GP and OUT_DC. *RPL32* was ranked second in OUT_GP and ALL but ninth in IND, and *UBCE* ranked seventh in OUT_DC and eighth in IND. Overall, with the BestKeeper algorithm *PGK*, *eIF2-1* and *KAS* were the least stable genes.

Comparative ΔC_T analysis. The comparative ΔC_T method uses the average standard deviation (SD) value of candidate reference genes as an indicator of expression stability. Genes with the lowest SD values are the most stable. As a whole, the highest stability of reference genes was found for IND samples and the 14 stable genes had an SD value < 1, compared to four reference genes for OUT_GP samples and one reference gene for OUT_DC samples (Fig 5). *EFL* ranked first in OUT_GP samples and *UBCE* was the most stable gene in OUT_DC and IND samples. *RPS10* had the lowest SD in the pooled samples, ALL, and was always among the top five most stable genes across all experimental conditions OUT_GP, OUT_DC and IND.

Comprehensive ranking of RefFinder. The web-based program RefFinder integrates the results of the four above-mentioned approaches (geNorm, NormFinder, BestKeeper and ΔC_T

Table 2. Stability values of candidate reference genes as determined by BestKeeper in the different culture conditions.

OUT_GP			OUT_DC			IND			ALL		
Gene symbol	SD	CV	Gene symbol	SD	CV	Gene symbol	SD	CV	Gene symbol	SD	CV
<i>UBCE</i>	0.48	2.11	<i>RPL32</i>	0.20	1.10	<i>EFL</i>	0.22	1.35	<i>EFL</i>	0.37	2.30
<i>RPL32</i>	0.53	2.91	<i>EFL</i>	0.25	1.57	<i>aTUB-1</i>	0.24	1.24	<i>RPL32</i>	0.40	2.24
<i>EFL</i>	0.56	3.34	<i>RPS10</i>	0.28	1.53	<i>ALD</i>	0.25	1.21	<i>UBCE</i>	0.53	2.29
<i>His2A</i>	0.59	2.62	<i>rbcL</i>	0.32	2.59	<i>eIF2-2</i>	0.33	1.32	<i>cdkA</i>	0.63	2.50
<i>rbcL</i>	0.60	4.76	<i>eIF2-2</i>	0.47	1.91	<i>aTUB-2</i>	0.33	1.44	<i>RPS10</i>	0.63	3.36
<i>cdkA</i>	0.65	2.54	<i>GAPDH</i>	0.47	1.86	<i>cdkA</i>	0.33	1.33	<i>rbcL</i>	0.65	5.08
<i>aTUB-2</i>	0.75	3.07	<i>UBCE</i>	0.62	2.73	<i>GAPDH</i>	0.37	1.53	<i>eIF2-2</i>	0.72	2.86
<i>bTUB</i>	0.75	3.37	<i>18S</i>	0.68	6.44	<i>UBCE</i>	0.43	1.84	<i>ALD</i>	0.73	3.41
<i>aTUB-1</i>	0.78	3.91	<i>ACT</i>	0.78	3.60	<i>RPL32</i>	0.44	2.45	<i>aTUB-1</i>	0.77	3.97
<i>ACT</i>	0.81	3.57	<i>His2A</i>	0.81	3.63	<i>eIF2-1</i>	0.44	1.66	<i>18S</i>	0.77	6.87
<i>eIF2-2</i>	0.83	3.15	<i>cdkA</i>	0.90	3.64	<i>His2A</i>	0.63	2.70	<i>ACT</i>	0.78	3.50
<i>18S</i>	0.85	7.63	<i>aTUB-2</i>	0.93	3.72	<i>RPS10</i>	0.68	3.60	<i>His2A</i>	0.78	3.45
<i>RPS10</i>	0.85	4.48	<i>ALD</i>	0.93	4.40	<i>ACT</i>	0.69	3.10	<i>aTUB-2</i>	0.94	3.91
<i>ALD</i>	0.89	4.04	<i>KAS</i>	1.03	4.51	<i>18S</i>	0.70	6.00	<i>bTUB</i>	0.98	4.44
<i>KAS</i>	0.95	4.05	<i>aTUB-1</i>	1.07	5.73	<i>bTUB</i>	0.82	3.67	<i>KAS</i>	1.15	4.96
<i>GAPDH</i>	1.35	5.02	<i>bTUB</i>	1.18	5.51	<i>rbcL</i>	0.82	6.19	<i>GAPDH</i>	1.19	4.68
<i>eIF2-1</i>	1.44	4.93	<i>eIF2-1</i>	1.65	6.16	<i>KAS</i>	1.51	6.57	<i>eIF2-1</i>	1.42	5.17
<i>PGK</i>	1.62	7.51	<i>PGK</i>	1.72	8.31	<i>PGK</i>	1.72	7.70	<i>PGK</i>	1.80	8.36

OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: Indoor cultures, ALL: Combination of the three previous conditions. SD: Standard deviation, CV: Coefficient of variation.

<https://doi.org/10.1371/journal.pone.0245495.t002>

method), and calculates the geometric mean of the ranking of all the candidate reference genes included in the analysis. In OUT_GP samples, the most stable genes were *EFL*, *ALD* and *ACT*, whereas *RPS10*, *rbcL* and *UBCE* were the most suitable reference genes in OUT_DC (Fig 6). In IND, *UBCE*, *aTUB-2* and *cdkA* exhibited the highest expression stability. And in pooled samples, ALL, *EFL*, *RPS10* and *RPL32* were the most stable genes.

In summary, quite different rankings of the best candidate reference genes were observed when experimental samples from different culture conditions were compared and analyzed using different methods, which highlights that the results can vary significantly with the algorithm used. The geNorm output indicated that two reference genes were enough for normalization in all the conditions tested. Comparison of the ranking of the candidate reference genes with the different methods, revealed that the optimum reference genes were *ALD* and *EFL* for OUT_GP, *RPL32* and *UBCE* for OUT_DC and *cdkA* and *UBCE* for IND. Moreover, *EFL* and *cdkA* or *EFL* and *UBCE* were the most appropriate for the pooled samples (ALL). In all the above recommendations, pairs of genes belonging to the same functional group were excluded to avoid possible existence of co-regulation that might interfere with the reliability of the results. For instance, although *EFL*, *RPS10* and *RPL32* were among the most stable genes for ALL, these three genes are involved in protein biosynthesis and hence only *EFL* (as top-ranked by geNorm, BestKeeper and RefFinder) was chosen.

Expression profile of AGPL and AGPS in OUT_DC

ADP-glucose pyrophosphorylase catalyzes the first step of starch biosynthesis, that is, the synthesis of ADP-glucose. In eukaryotes, it is a heterotetramer comprised of two distinct subunits, two alpha (AGPS) and two beta (AGPL). As starch accumulates due to carbon fixation during

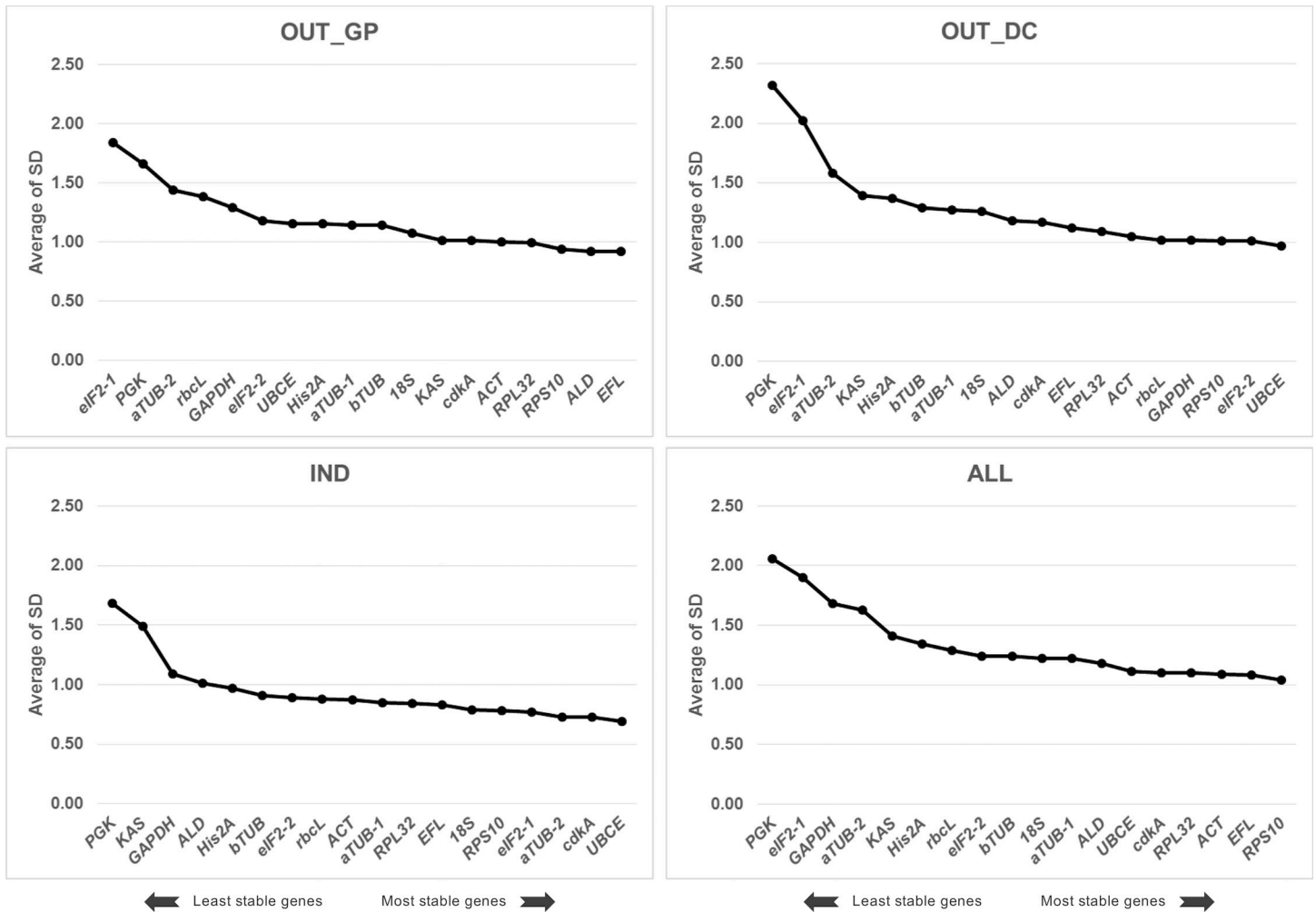


Fig 5. Global ranking of the 18 candidate reference genes using the comparative ΔC_t method. The expression stability value was determined in all tested samples (a total of 45) and conditions (OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: Indoor cultures, ALL: Combination of the three previous conditions). A lower value indicates a more stable expression pattern.

<https://doi.org/10.1371/journal.pone.0245495.g005>

photosynthesis and is further degraded in the dark, we hypothesized that *AGPS* and *AGPL* were likely co-regulated and would have similar expression patterns modulated by starch content fluctuations. To test this hypothesis and verify the validated reference genes in RT-qPCR, we quantified *AGPL* and *AGPS* expression in OUT_DC samples. The reference genes for normalization were *RPL32* and *UBCE*, since they were the most stable for the OUT_DC samples and this was compared with the outcome of normalization with the least stable genes *PGK* and *eIF2-1*. In addition, *AGPL* and *AGPS* expression was normalized with *RPS10* and *rbcL*, the most stable reference genes for OUT_DC samples using the RefFinder program ranking. As a complementary analysis and considering the results obtained in the combined condition ALL, the suitability of other candidate reference genes such as *cdkA* and *ACT* in combination with *EFL* was also evaluated. As shown in Fig 7, when *RPL32* and *UBCE* genes were used for normalization, either alone or in combination, a statistically significant increase in *AGPL* transcripts was observed at 7 p.m. being 2.67-fold (*UBCE*) and 3.23-fold (*RPL32*) higher than at 7 a.m. (the calibrator time point). The gene expression results obtained with *rbcL*, *RPS10* and their combination matched those previously mentioned, with a peak in transcript abundance

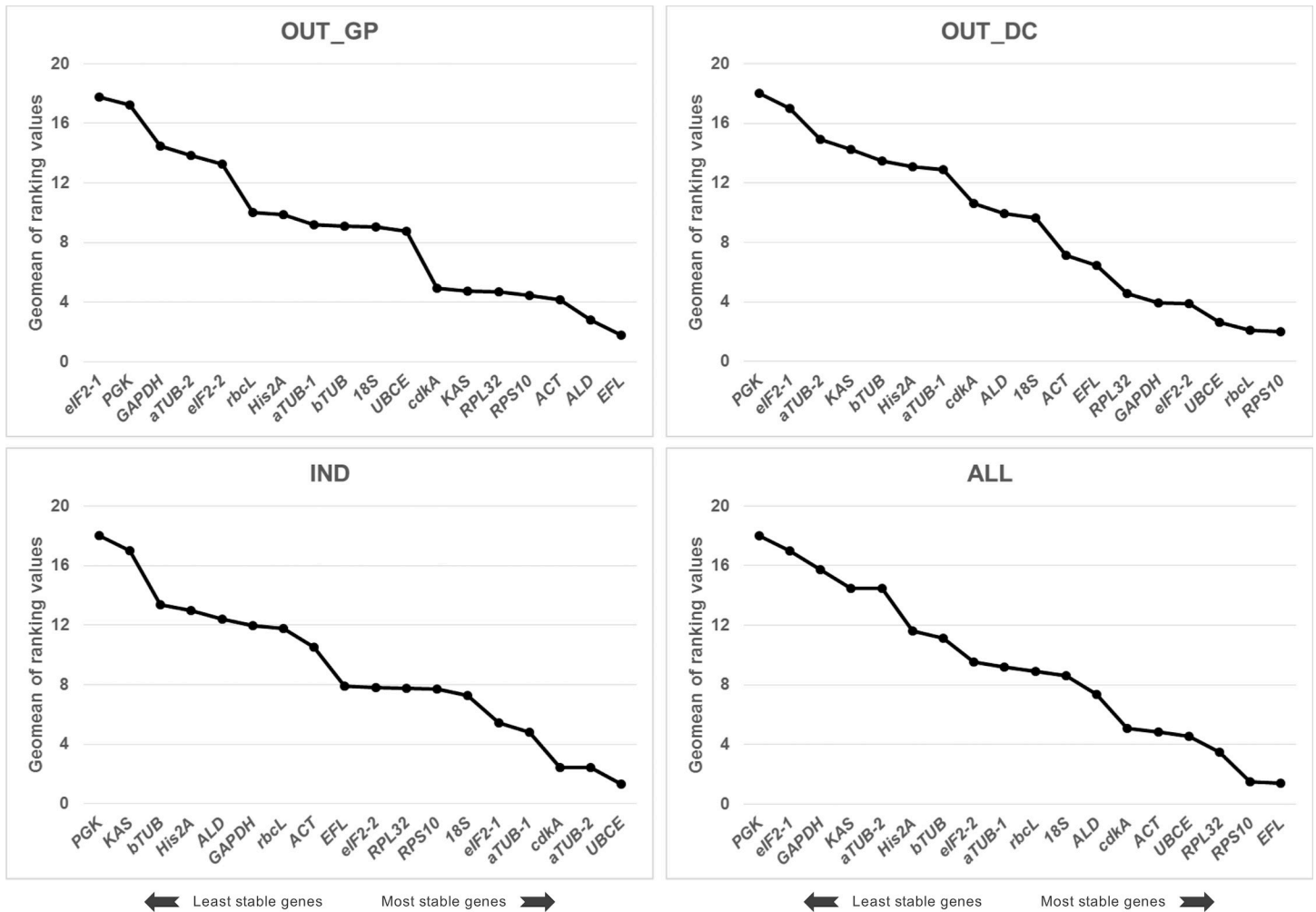


Fig 6. Comprehensive ranking of the 18 candidate reference genes using RefFinder. The expression stability value was determined in all tested samples (a total of 45) and conditions (OUT_GP: Outdoor cultures during the growth phase, OUT_DC: Outdoor cultures during the semi-continuous phase, IND: Indoor cultures, ALL: Combination of the three previous conditions). A lower value indicates a more stable expression pattern.

<https://doi.org/10.1371/journal.pone.0245495.g006>

at 7 p.m. (ranging between 2.76-fold and 3.10-fold higher than at 7:00 a.m. with *rbcl* and *RPS10*, respectively). This profile was also conserved when the data were analyzed with *EFL* and *cdkA*, or *EFL* and *ACT* (S6 Fig), with some slight differences in the ratios. However, when the analysis was performed with the least stable genes *PGK* and *eIF2-1*, strong differences could be found. Although the expression pattern with *PGK* was similar to those found with stable genes in the sense that a peak in gene expression was detected at 7 p.m. and the lowest transcript levels were found at 10 a.m., the ratios were significantly different. Up to 90.81-fold higher transcripts were measured at 7 p.m., representing a higher than 500 fold-increase in relation to 10 a.m., this ratio being significantly higher than the ratios found with the most stable genes (always lower than 25-fold). Moreover, when *eIF2-2* was employed as internal control, it was striking that the highest expression levels of *AGPL* were detected at 4 p.m. (~28-fold higher than at 7 a.m.), although the lowest transcript amounts were measured again at 10 a.m. (0.48-fold lower than at 7 a.m.). With the combination of *PGK* and *eIF2-1* the maximum and minimum expression levels were observed again at 7 p.m. and 10 a.m. respectively, but with ratios not only at 7 p.m. but also at 1 p.m. and 4 p.m. being strongly different to those

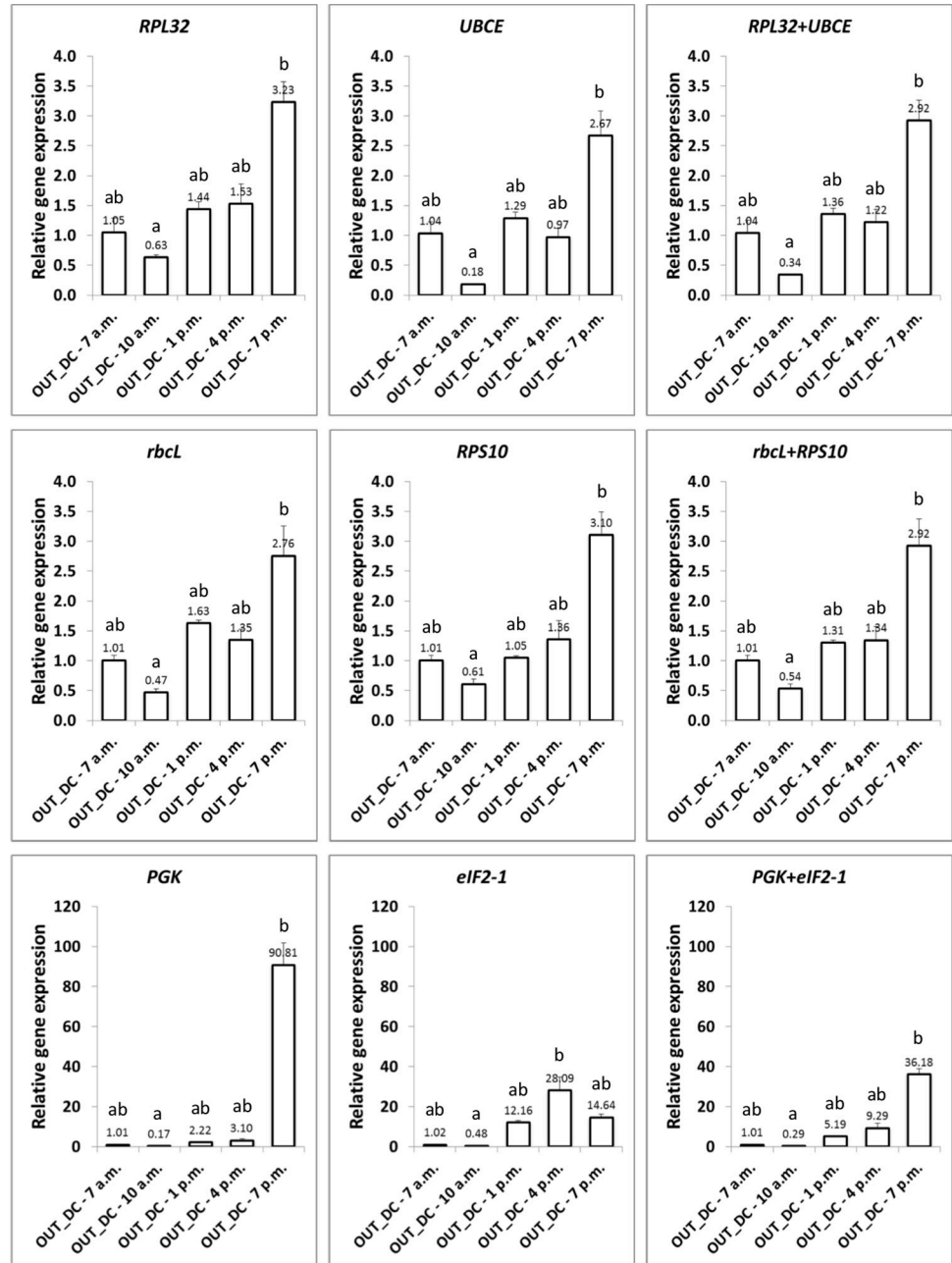


Fig 7. Relative AGPL expression profiles in OUT_DC samples determined by qRT-PCR using different reference genes. In all instances, data are expressed (at the outside end of the corresponding column) as the mean fold change (mean + SEM, n = 3) from the calibrator group (OUT_DC—7 a.m.). Different letters denote significant differences ($P < 0.05$) between time points using the Friedman test (non-parametric one-way ANOVA) followed by the Dunn’s multiple comparison test.

<https://doi.org/10.1371/journal.pone.0245495.g007>

shown with the most stable genes. As a whole, analyses with stable reference genes fitted our hypothesis in the sense that the highest transcript amounts of AGPL in OUT_DC were observed at the end of the light phase, when maximum starch accumulation might be expected.

The expression of AGPS was evaluated using the same reference genes for normalization as was used for AGPL (Fig 8). The AGPS gene expression profile and expression ratios after

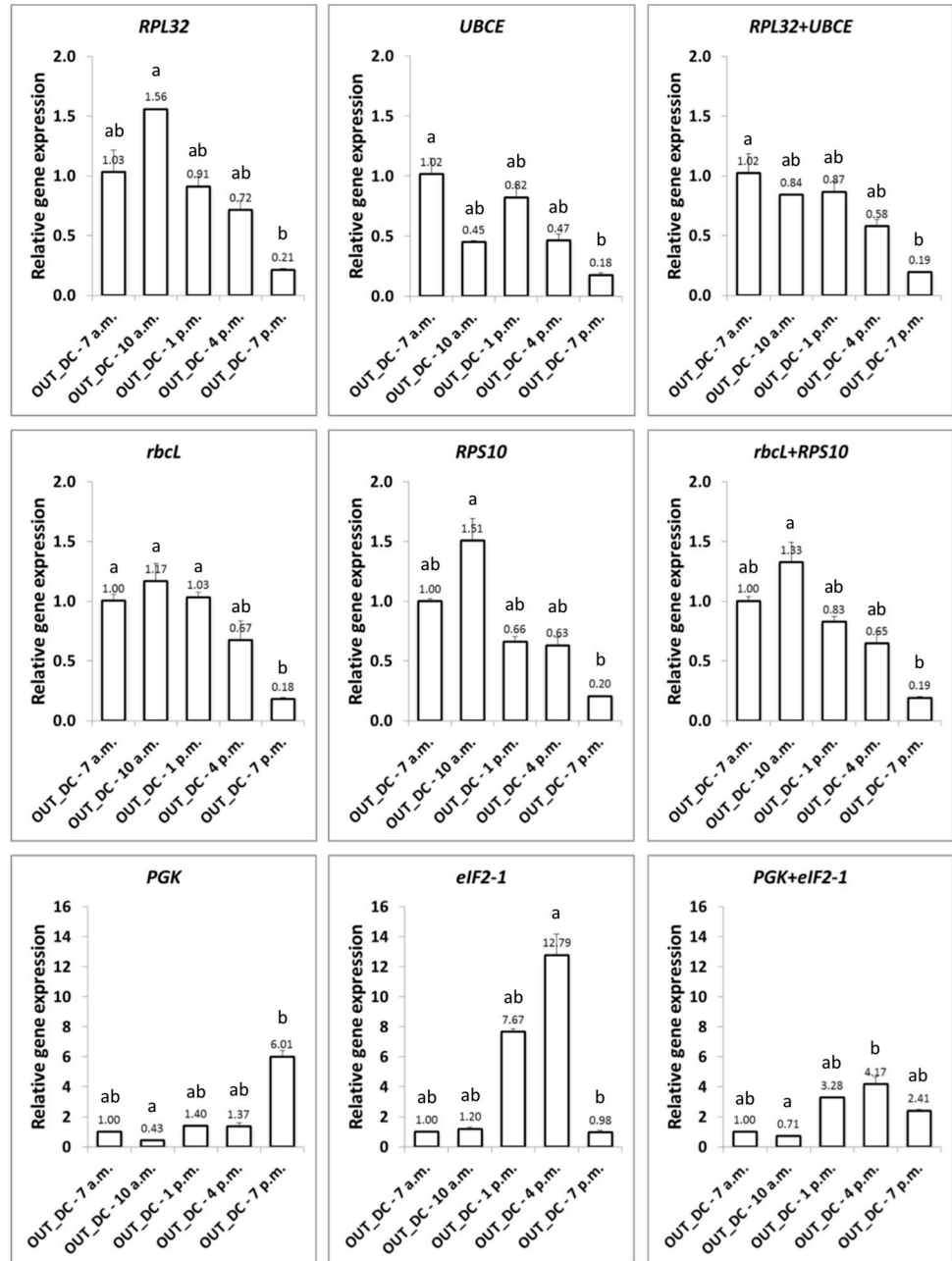


Fig 8. Relative AGPS expression profiles in OUT_DC samples determined by qRT-PCR using different reference genes. In all instances, data are expressed (at the outside end of the corresponding column) as the mean fold change (mean + SEM, n = 3) from the calibrator group (OUT_DC—7 a.m.). Different letters denote significant differences ($P < 0.05$) between times points using the Friedman test (non-parametric one-way ANOVA) followed by the Dunn’s multiple comparison test.

<https://doi.org/10.1371/journal.pone.0245495.g008>

normalization with *RPL32* and *UBCE* alone or in combination, revealed a significant down-regulation at 7 p.m. (~5-fold lower transcripts) compared to 7 a.m. (used as a calibrator). Similar results were also obtained for AGPS expression normalized with *rbcl*, *RPS10* and their combination. The large drop in AGPS transcription at 7:00 p.m. was even more pronounced (~6-fold lower than at 7:00 a.m.) when *EFL* and *cdkA* or *EFL* and *ACT* were used for

normalization (S6 Fig). However, expression patterns were substantially modified when normalization was performed using the least stable reference genes. For example, a significant up-regulation of *AGPS* was observed at 7 p.m. when *PGK* was used for normalization (6.01-fold higher than at 7 a.m.), and at 4 p.m. when *eIF2-1* was used for normalization (12.79-fold higher) or when both genes were used (4.17-fold higher). In contrast to what was initially hypothesized, transcriptional regulation of *AGPS* was opposite to that of *AGPL*, as the lowest transcript abundance of *AGPS* occurred concomitantly with the highest transcription of *AGPL* at the end of the light phase.

In conclusion, when appropriate reference genes are selected for normalization in gene expression analysis using different algorithms similar outcomes are obtained. However, inappropriate selection of reference genes is a major concern as it can lead to substantially modified gene expression patterns.

Discussion

In this study, the expression stability of 18 candidate reference genes was evaluated in the green microalgae species *T. chui* using geNorm, NormFinder, BestKeeper, the comparative ΔC_T method, and RefFinder. Three different experimental conditions for microalgae production were tested (OUT_GP, OUT_DC, and IND), as well as a pool of all samples (ALL). The most prominent observation was that each of the above tools produced a different set of top-ranked reference genes in each group of experimental samples and conditions. This finding was not entirely unexpected, as such heterogeneous results may be caused by application of different mathematical algorithms as found in previous studies [21,33,35,37]. As each algorithm has its own drawbacks [63], the use of different methods to define the most appropriate reference genes for target gene expression normalization in RT-qPCR analysis is strongly recommended to improve the reliability of results. Moreover, to our knowledge, the present study represents the first RT-qPCR survey of candidate genes for normalization in *T. chui*. The only related study was focused on the expression of a high-affinity phosphate transporter gene in *T. chui* using absolute rather than relative PCR quantification so that no normalization was needed [64]. Absolute quantification requires construction of a standard sense RNA for every target gene of interest to create an external calibration curve, and this other well-known constraints of absolute quantification [65], makes it significantly less practical when quantifying ten or more genes in a single study. Herein the suitability of 18 candidate reference genes for relative real-time expression analysis in microalgae samples harvested in a production setting from 4000 L photobioreactors under different conditions, OUT_GP and OUT_DC, was evaluated. The main stress conditions for microalgae were related to the daily environmental changes, with light intensity and temperature being probably the most relevant factors. This study aimed to have a final application in industry and as such is very different from the highly controlled, low culture volumes of laboratory conditions, either standard or stressful, previously used to identify reference for other microalgae species [44,47,49–53,66].

Gene expression analyses represent a powerful tool for gaining insight into the molecular and biochemical mechanisms that govern physiological features and responses to environmental changes in a wide range of different organisms, including microalgae. For this reason high throughput sequencing approaches have been employed to provide a comprehensive view of the responsiveness of different metabolic pathways under defined experimental conditions, nitrogen starvation being probably the most widely assessed [58,67–69]. Once the genetic resources containing thousands of predicted transcripts are generated and become available, such information can be further processed and used to study the expression of selected sets of genes involved in relevant metabolic pathways through RT-qPCR platforms containing a

variable number of representative genes. For instance, 16 differentially expressed genes involved in lipid accumulation and carbon fixation were studied in the species *Chorella sorokiniana* grown under nitrogen limiting conditions to validate RNA-seq data [70]. In *Tetraselmis suecica*, a set of 14 selected genes exhibited significant up- and down-regulation using RT-qPCR in agreement with RNA-seq results [58]. Recently, a RT-qPCR platform containing more than 100 targeted genes was used for *Nannochloropsis gaditana* to study the influence of light quality on metabolism [71]. A similar strategy with an array of ~100 genes was also employed in human cell lines to screen for the potential bioactivity of microalgae extracts [72]. These studies demonstrate that modulation of metabolic processes can be detected via gene expression analysis of gene markers. When dealing with microalgae production at an industrial scale, expression analysis with a species-specific RT-qPCR platform containing selected gene markers represents an innovative biotechnological approach for quality control of biomass. For *T. chui*, although there are available genetic resources, biotechnological tools still need to be developed. The present study identifying appropriate reference genes represents a first and key step for the accurate detection of varying gene expression under different conditions.

Most of the candidate reference genes included in this study have previously been tested in other microalgae species. In particular, the stability of the traditional *GAPDH*, *ACT*, and *18S* reference genes has been simultaneously assessed in a range of microalgae species [17,43,47,49,50,52,53,55,70], macroalgae [40,73,74] and plants [22,75,76]. The two candidate alpha tubulin encoding genes here analyzed, *aTUB-1* and *aTUB-2*, probably arose by gene duplication and this observation is not restricted to *T. chui*. Tubulin belongs to a multigene family in other eukaryotes including mammals, ciliates, fungi, or plants [77–80], and a variable number of gene duplicates have also been found in other evolutionary distant microalgae such as *Bigeloviella*, *Cyanophora*, *Emiliania* or *Chlamydomonas* [81]. Two genes encoding the eukaryotic translation initiation factor 2 (*eIF2-1* and *eIF2-2*) were identified in *T. chui* and were analyzed. The *eIF2-2* isoform in *T. chui* was most similar to mitochondrial isoforms, but lacked a signal peptide as shown by analysis with SignalP 5.0 [82]. To our knowledge, this is the first time that these gene duplicates have been tested as reference genes in microalgae, as the eukaryotic translation initiation factor 4E (*eIF4E*) was assessed in the dinoflagellates *Prorocentrum minimum* [47] and *Karenia mikimotoi* [50], and another gene referred to as eukaryotic (translation) initiation factor was tested in the green microalgae *Closterium ehrenbergii* [56] or the dinoflagellate *Alexandrium catanella* [49].

The *EFL* gene included in this study is related to, but clearly distinguishable from the eukaryotic elongation factor-1 alpha gene (*EF-1 α*) previously analyzed in several microalgae [52–54]. *EF-1 α* expression was also assessed in green microalgae such as *Volvox carteri* [55] and *Chlamydomonas* [17,57], although since only *EFL* has been reported in green algae it seems likely that the putative *EF-1 α* was misannotated [66,83]. In addition, the inclusion in the present study of *KAS* as a potential reference gene for normalization is new to microalgae, and came from previous data about gene stability in the microalgae *Nannochloropsis gaditana* [84]. Finally, the remaining genes (*bTUB*, *PGK*, *His2A*, *rbcL*, *cdkA*, *UBCE*, *ALD*, *RPL32* and *RPS10*) were selected as they have been previously tested as reference genes in microalgae.

The results for the candidate reference genes in *T. chui* compared to previous studies of other microalgae species revealed they were variable. For instance, although *GAPDH* and *ACT* are the most used reference genes for normalization in metazoan [26], few studies have demonstrated them to be optimal for normalization. For example, *GAPDH* was the most stable gene under different experimental conditions in *Chlamydomonas* [17,57], but in the present study it was not top-ranked. In contrast, *ACT*, which led the stability ranking in *Chlamydomonas* sp. [17], *Nannochloropsis* sp. [52], *Ditylum brightwellii* [45], *Closterium ehrenbergii* [56],

Isochrysis zhangjiangensis [51], *Tetraselmis suecica* [58], and was the second most stable gene in *Alexandrium catanella* [49], was not well ranked in our study or *Volvox* [55]. Similarly, *18S* was not the most stable gene in any condition in *T. chui*.

The use of alpha and beta tubulin genes has been reported in other microalgae species and in some cases these genes were two of the most stable genes [44]. In other studies only one of the two genes was stable [47,50–52,54,56,58], and in other microalgae species none of the two genes was among those with the highest stability [17,43,55,57]. In this study, *bTUB* was never among the top seven most stable genes in any of the analysis or conditions assayed, and with the exception of IND, neither *aTUB-1* nor *aTUB-2* was in the first seven ranked genes. Overall, our study revealed that choice of the reference gene depends on species, experimental conditions and the analytical approach, highlighting the need for appropriate stability testing of reference genes before their use in expression analysis of target genes.

From an industrial perspective we assessed the expression of the *T. chui* genes encoding both the large (*AGPL*) and small (*AGPS*) subunits of ADP-glucose pyrophosphorylase, as the functional enzyme is a hetero-tetrameric protein composed of two alpha (or small, S) and two beta (or large, L) subunits [85,86]. The results confirmed that the use of inappropriate reference genes can lead to strong differences not only in the expression profile of the target gene, but also in its relative expression levels (expressed as fold-changes) between sampling points. This is in agreement with the findings of previous studies [17,52,54,87], which highlights the importance of validating reference gene stability. The second main objective was to study possible transcriptional regulation of both genes in an outdoor production system. The ADP-glucose pyrophosphorylase is involved in the synthesis of ADP-glucose, which is the donor for the elongation of alpha-1,4-glucosidic chains that serve as intracellular carbon and energy storage, as starch in plants and green and red algae and glaucophytes [88,89]. In microalgae, starch production can be strongly affected by light intensity and temperature [90,91], factors that significantly varied in samples collected from OUT_DC. Although *AGPL* and *AGPS* are both involved in starch synthesis the genes had a different regulation, as the peak of *AGPL* expression by the end of the light phase coincided with the lowest expression levels of *AGPS*. In a diurnal cycle, an increase in transcripts of ADP-glucose pyrophosphorylase gene was reported for the picoalgae *Ostreococcus* [92]. This finding suggests that *AGPL* and *AGPS* have different daily rhythms in *T. chui* under natural photoperiod. In this regard, genes exhibiting differential transcriptional regulation under light/dark cycles have been reported in a range of different microalgae such as *Chlamydomonas* [93,94], *Ostreococcus* [93,95], *Nannochloropsis* [96], *Phaeodactylum* [97], *Tetrademus* [98], *Bigeloviella* [99], or *Chrysochromulina* [100]. More research will be needed to confirm if the transcriptional pattern of *AGPL* and *AGPS* is conserved in a diel cycle, as well as to analyze the effect of changing climate conditions on their expression.

Conclusions

As far as we are aware the present study represents the first systematic analysis of candidate reference genes for the species *T. chui*. Moreover, and in contrast to all previous reports on microalgae, the samples analyzed were collected from large scale outdoor photobioreactors from operating industrial production units. For each of the conditions assessed (OUT_GP, OUT_DC, and IND), distinct rankings of genes were obtained with the tested algorithms. Overall, the following reference genes were identified as stably-expressed: *ALD* and *EFL* for OUT_GP, *RPL32* and *UBCE* for OUT_DC, and *cdkA* and *UBCE* for IND. Moreover, the genes *EFL* and *cdkA* or *EFL* and *UBCE* were the most appropriate for pooled samples (ALL). Expression analysis of the genes *AGPL* and *AGPS* in OUT_DC samples confirmed that the selected

reference genes were appropriate for normalization of target genes and the consequence of inappropriate reference genes for accuracy and reliability of the quantification. Finally, the present study forms the basis of an RT-qPCR platform to be used for quality control of *T. chui* biomass in industrial production facilities.

Supporting information

S1 Fig. Evolution of cell density in indoor cultures. The average \pm SD of three 5 L flasks is represented. Samples were collected after inoculation on day 1 (D1), and then at D4, D7, D10 and D13.

(TIF)

S2 Fig. Evolution of cell density in outdoor cultures. The average \pm SD of three 4000 L photobioreactors of the same production unit is represented. Samples were collected after inoculation on day 1 (D1), and then at D4, D7, D11 and D18.

(TIF)

S3 Fig. Agarose gel electrophoresis of PCR amplified products. The expected size of each amplicon (in bp) is shown. The 25 bp DNA Ladder (Invitrogen) was used as the molecular weight standard and the size (in bp) of the most relevant bands are shown. The black dashed line in the gel of the lower panel (TARGET GENES) indicates where irrelevant lanes were removed.

(TIF)

S4 Fig. Standard curves generated by all primer pairs. The curves were used for the determination of PCR efficiencies. Serial dilutions of input cDNA were plotted against the C_T values obtained by real-time PCR.

(TIF)

S5 Fig. Melting curves generated by all primer pairs used in the study. The melt curves were captured after cycle 40 by heating from 70°C to 95°C with a ramp speed of 0.5 s every 10 s.

(TIF)

S6 Fig. Relative *AGPL* (upper panel) and *AGPS* (lower panel) expression profiles in OUT_DC samples. The expression of *AGPL* and *AGPS* were determined by RT-qPCR using different reference genes. In all instances, data are expressed (at the outside end of the corresponding column) as the mean fold change (mean + SEM, $n = 3$) from the calibrator group (OUT_DC—7 a.m.). Different letters denote significant differences ($P < 0.05$) between time points using the Friedman test (non-parametric one-way ANOVA) followed by the Dunn's multiple comparison test.

(TIF)

S1 Table. Culture temperature (average \pm SD, °C) and PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$) figures at sampling days and times in outdoor conditions.

(DOCX)

S2 Table. Candidate reference genes for normalization of RT-qPCR expression data. Information regarding the target genes *AGPL* and *AGPS* included in the study is also shown below.

(DOCX)

S3 Table. Mean C_T values of all samples included in this study. Data are presented for each of the candidate reference genes.

(XLSX)

S1 Raw images.
(PDF)

Acknowledgments

We thank Deborah M. Power at Universidade do Algarve (Portugal) for critical review and editing of the manuscript.

Author Contributions

Conceptualization: Emmanouil Flemetakis, Carlos Infante.

Formal analysis: Emmanouil Flemetakis, Carlos Infante.

Funding acquisition: Lalia Mantecón.

Investigation: Sonia Torres, Carmen Lama.

Project administration: Lalia Mantecón.

Supervision: Lalia Mantecón.

Visualization: Carlos Infante.

Writing – original draft: Carlos Infante.

Writing – review & editing: Lalia Mantecón, Emmanouil Flemetakis.

References

1. Tirichine L, Bowler C. Decoding algal genomes: Tracing back the history of photosynthetic life on Earth. *Plant J.* 2011; 66: 45–57. <https://doi.org/10.1111/j.1365-313X.2011.04540.x> PMID: 21443622
2. Heimann K, Huerlimann R. Microalgal Classification: Major Classe sand Genera of Commercial Microalgal Species. *Handbook of Marine Microalgae: Biotechnology Advances.* Elsevier Inc.; 2015. pp. 25–41. <https://doi.org/10.1016/B978-0-12-800776-1.00003-0>
3. Bule MH, Ahmed I, Maqbool F, Bilal M, Iqbal HMN. Microalgae as a source of high-value bioactive compounds. *Front Biosci—Sch.* 2018; 10: 197–216. <https://doi.org/10.2741/s509> PMID: 28930527
4. Vaz B da S, Moreira JB, Morais MG de, Costa JAV. Microalgae as a new source of bioactive compounds in food supplements. *Curr Opin Food Sci.* 2016; 7: 73–77. <https://doi.org/10.1016/j.cofs.2015.12.006>
5. García JL, de Vicente M, Galán B. Microalgae, old sustainable food and fashion nutraceuticals. *Microb Biotechnol.* 2017; 10: 1017–1024. <https://doi.org/10.1111/1751-7915.12800> PMID: 28809450
6. Fu W, Nelson DR, Yi Z, Xu M, Khraiwesh B, Jijakli K, et al. Bioactive Compounds From Microalgae: Current Development and Prospects. *Stud Nat Prod Chem.* 2017; 54: 199–225. <https://doi.org/10.1016/B978-0-444-63929-5.00006-1>
7. Barkia I, Saari N, Manning SR. Microalgae for high-value products towards human health and nutrition. *Mar Drugs.* 2019; 17: 304. <https://doi.org/10.3390/md17050304> PMID: 31137657
8. Borowitzka MA. *Biology of Microalgae. Microalgae in Health and Disease Prevention.* Elsevier Inc.; 2018. <https://doi.org/10.1016/b978-0-12-811405-6.00003-7>
9. Conceição LEC, Yúfera M, Makridis P, Morais S, Dinis MT. Live feeds for early stages of fish rearing. *Aquac Res.* 2010; 41: 613–640. <https://doi.org/10.1111/j.1365-2109.2009.02242.x>
10. Brown MR, Jeffrey SW. Biochemical composition of microalgae from the green algal classes Chlorophyceae and Prasinophyceae. 1. Amino acids, sugars and pigments. *J Exp Mar Bio Ecol.* 1992; 161: 91–113. [https://doi.org/10.1016/0022-0981\(92\)90192-D](https://doi.org/10.1016/0022-0981(92)90192-D)
11. Tibbetts SM, Milley JE, Lall SP. Chemical composition and nutritional properties of freshwater and marine microalgal biomass cultured in photobioreactors. *J Appl Phycol.* 2015; 27: 1109–1119. <https://doi.org/10.1007/s10811-014-0428-x>
12. Rahman NA, Khatoon H, Yusuf N, Banerjee S, Haris NA, Lananan F, et al. *Tetraselmis chui* biomass as a potential feed additive to improve survival and oxidative stress status of Pacific white-leg shrimp

- Litopenaeus vannamei postlarvae. *Int Aquat Res.* 2017; 9: 235–247. <https://doi.org/10.1007/s40071-017-0173-2>
13. Cerezuela R, Guardiola FA, Meseguer J, Esteban MÁ. Enrichment of gilthead seabream (*Sparus aurata* L.) diet with microalgae: Effects on the immune system. *Fish Physiol Biochem.* 2012; 38: 1729–1739. <https://doi.org/10.1007/s10695-012-9670-9> PMID: 22678706
 14. Cerezuela R, Guardiola FA, González P, Meseguer J, Esteban MÁ. Effects of dietary *Bacillus subtilis*, *Tetraselmis chuii*, and *Phaeodactylum tricornutum*, singularly or in combination, on the immune response and disease resistance of sea bream (*Sparus aurata* L.). *Fish Shellfish Immunol.* 2012; 33: 342–349. <https://doi.org/10.1016/j.fsi.2012.05.004> PMID: 22634255
 15. Mantecón L, Moyano R, Cameán AM, Jos A. Safety assessment of a lyophilized biomass of *Tetraselmis chuii* (TetraSOD®) in a 90 day feeding study. *Food Chem Toxicol.* 2019; 133: 110810. <https://doi.org/10.1016/j.fct.2019.110810> PMID: 31505236
 16. Hu R, Fan C, Li H, Zhang Q, Fu YF. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol Biol.* 2009; 10: 93. <https://doi.org/10.1186/1471-2199-10-93> PMID: 19785741
 17. Mou S, Zhang X, Miao J, Zheng Z, Xu D, Ye N. Reference genes for gene expression normalization in *Chlamydomonas* sp. ICE-L by quantitative real-time RT-PCR. *J Plant Biochem Biotechnol.* 2015; 24: 276–282. <https://doi.org/10.1007/s13562-014-0268-4>
 18. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques.* 2008; 44: 619–626. <https://doi.org/10.2144/000112776> PMID: 18474036
 19. Wang Z, Meng Q, Zhu X, Sun S, Gao S, Gou Y, et al. Evaluation and Validation of Reference Genes for Quantitative Real-Time PCR in *Helopeltis theivora* Waterhouse (Hemiptera: Miridae). *Sci Rep.* 2019; 9: 13291. <https://doi.org/10.1038/s41598-019-49479-1> PMID: 31527603
 20. Köhler M, Leitsch D, Müller N, Walochnik J. Validation of reference genes for the normalization of RT-qPCR gene expression in *Acanthamoeba* spp. *Sci Rep.* 2020; 10: 10362. <https://doi.org/10.1038/s41598-020-67035-0> PMID: 32587282
 21. Sarker N, Fabijan J, Emes RD, Hemmatzadeh F, Meers J, Moreton J, et al. Identification of stable reference genes for quantitative PCR in koalas. *Sci Rep.* 2018; 8: 3364. <https://doi.org/10.1038/s41598-018-21723-0> PMID: 29463845
 22. Wang J, Abbas M, Wen Y, Niu D, Wang L, Sun Y, et al. Selection and validation of reference genes for quantitative gene expression analyses in black locust (*Robinia pseudoacacia* L.) using real-time quantitative PCR. *PLoS One.* 2018; 13: e0193076. <https://doi.org/10.1371/journal.pone.0193076> PMID: 29529054
 23. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol Genet Genomics.* 2010; 283: 233–241. <https://doi.org/10.1007/s00438-010-0511-1> PMID: 20098998
 24. Kozera B, Rapacz M. Reference genes in real-time PCR. *J Appl Genet.* 2013; 54: 391–406. <https://doi.org/10.1007/s13353-013-0173-x> PMID: 24078518
 25. Guénin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L. Normalization of qRT-PCR data: The necessity of adopting a systematic, experimental conditions-specific, validation of references. *J Exp Bot.* 2009; 60: 487–493. <https://doi.org/10.1093/jxb/ern305> PMID: 19264760
 26. Chapman JR, Waldenström J. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. *PLoS One.* 2015; 10: e0141853. <https://doi.org/10.1371/journal.pone.0141853> PMID: 26555275
 27. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002; 3: research0034. <https://doi.org/10.1186/gb-2002-3-7-research0034> PMID: 12184808
 28. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 2004; 64: 5245–5250. <https://doi.org/10.1158/0008-5472.CAN-04-0496> PMID: 15289330
 29. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol Lett.* 2004; 26: 509–515. <https://doi.org/10.1023/b:bile.0000019559.84305.47> PMID: 15127793
 30. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol.* 2006; 7: 33. <https://doi.org/10.1186/1471-2199-7-33> PMID: 17026756

31. Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol Biol*. 2012; 80: 75–84. <https://doi.org/10.1007/s11103-012-9885-2> PMID: 22290409
32. Schulze F, Malhan D, El Khassawna T, Heiss C, Seckinger A, Hose D, et al. A tissue-based approach to selection of reference genes for quantitative real-time PCR in a sheep osteoporosis model. *BMC Genomics*. 2017; 18: 975. <https://doi.org/10.1186/s12864-017-4356-4> PMID: 29258442
33. Aggarwal A, Jamwal M, Viswanathan GK, Sharma P, Sachdeva MUS, Bansal D, et al. Optimal Reference Gene Selection for Expression Studies in Human Reticulocytes. *J Mol Diagnostics*. 2018; 20: 326–333. <https://doi.org/10.1016/j.jmoldx.2018.01.009> PMID: 29474985
34. Aminfar Z, Rabiei B, Tohidfar M, Mirjalili MH. Selection and validation of reference genes for quantitative real-time PCR in *Rosmarinus officinalis* L. in various tissues and under elicitation. *Biocatal Agric Biotechnol*. 2019; 20: 101246. <https://doi.org/10.1016/j.bcab.2019.101246>
35. Kumar D, Das PK, Sarmah BK. Reference gene validation for normalization of RT-qPCR assay associated with germination and survival of rice under hypoxic condition. *J Appl Genet*. 2018; 59: 419–430. <https://doi.org/10.1007/s13353-018-0466-1> PMID: 30232619
36. Tang X, Zhang N, Si H, Calderón-Urrea A. Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods*. 2017; 13: 85. <https://doi.org/10.1186/s13007-017-0238-7> PMID: 29075311
37. Zhang J, Xie W, Yu X, Zhang Z, Zhao Y, Wang N, et al. Selection of suitable reference genes for rt-qpcr gene expression analysis in siberian wild rye (*Elymus sibiricus*) under different experimental conditions. *Genes (Basel)*. 2019; 10: 451. <https://doi.org/10.3390/genes10060451> PMID: 31200580
38. Zarivi O, Cesare P, Ragnelli AM, Aimola P, Leonardi M, Bonfigli A, et al. Validation of reference genes for quantitative real-time PCR in Périgord black truffle (*Tuber melanosporum*) developmental stages. *Phytochemistry*. 2015; 116: 78–86. <https://doi.org/10.1016/j.phytochem.2015.02.024> PMID: 25778998
39. Chen C, Xie T, Ye S, Jensen AB, Eilenberg J. Selection of reference genes for expression analysis in the entomophthoralean fungus *Pandora neoaphidis*. *Brazilian J Microbiol*. 2016; 47: 259–265. <https://doi.org/10.1016/j.bjm.2015.11.031> PMID: 26887253
40. Dong M, Zhang X, Chi X, Mou S, Xu J, Xu D, et al. The validity of a reference gene is highly dependent on the experimental conditions in green alga *Ulva linza*. *Curr Genet*. 2012; 58: 13–20. <https://doi.org/10.1007/s00294-011-0361-3> PMID: 22205301
41. Kong F, Cao M, Sun P, Liu W, Mao Y. Selection of reference genes for gene expression normalization in *Pyropia yezoensis* using quantitative real-time PCR. *J Appl Phycol*. 2015; 27: 1003–1010. <https://doi.org/10.1007/s10811-014-0359-6>
42. Shim J, Shim E, Kim GH, Han JW, Zuccarello GC. Keeping house: Evaluation of housekeeping genes for real-time PCR in the red alga, *Bostrychia moritziana* (Florideophyceae). *Algae*. 2016; 31: 167–174. <https://doi.org/10.4490/algae.2016.31.5.25>
43. Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, et al. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene*. 2007; 406: 23–35. <https://doi.org/10.1016/j.gene.2007.05.022> PMID: 17658702
44. Adelfi MG, Borra M, Sanges R, Montesor M, Fontana A, Ferrante MI. Selection and validation of reference genes for qPCR analysis in the pennate diatoms *Pseudo-nitzschia multistriata* and *P. arenysensis*. *J Exp Mar Bio Ecol*. 2014; 451: 74–81. <https://doi.org/10.1016/j.jembe.2013.11.003>
45. Guo R, Lee MA, Ki JS. Normalization genes for mRNA expression in the marine diatom *Ditylum brightwellii* following exposure to thermal and toxic chemical stresses. *J Appl Phycol*. 2013; 25: 1101–1109. <https://doi.org/10.1007/s10811-012-9908-z>
46. Liu Q, Xing Y, Li Y, Wang H, Mi T, Zhen Y, et al. Carbon fixation gene expression in *Skeletonema marinoi* in nitrogen-, phosphate-, silicate-starvation, and low-temperature stress exposure. *Journal of Phycology*. 2020. <https://doi.org/10.1111/jpy.12936> PMID: 31628865
47. Guo R, Ki JS. Evaluation and validation of internal control genes for studying gene expression in the dinoflagellate *Prorocentrum minimum* using real-time PCR. *Eur J Protistol*. 2012; 48: 199–206. <https://doi.org/10.1016/j.ejop.2011.11.001> PMID: 22209541
48. Rosic NN, Pernice M, Rodriguez-Lanetty M, Hoegh-Guldberg O. Validation of Housekeeping Genes for Gene Expression Studies in *Symbiodinium* Exposed to Thermal and Light Stress. *Mar Biotechnol*. 2011; 13: 355–365. <https://doi.org/10.1007/s10126-010-9308-9> PMID: 20668900
49. Niaz Z, Sui Z, Riaz S, Liu Y, Shang E, Xing Q, et al. Identification of valid reference genes for the normalization of RT-qPCR gene expression data in *Alexandrium catenella* under different nutritional conditions. *J Appl Phycol*. 2019; 31: 1819–1833. <https://doi.org/10.1007/s10811-018-1664-2>

50. Shi X, Xiao Y, Liu L, Xie Y, Ma R, Chen J. Exploring reliable reference genes for gene expression normalization in *Karenia mikimotoi* using real-time PCR. *J Appl Phycol*. 2020; 32: 431–440. <https://doi.org/10.1007/s10811-019-01961-z>
51. Wu S, Zhou J, Cao X, Xue S. Determination of internal controls for quantitative gene expression of *Isochrysis zhangjiangensis* at nitrogen stress condition. *J Ocean Univ China*. 2016; 15: 137–144. <https://doi.org/10.1007/s11802-016-2847-6>
52. Cao S, Zhang X, Ye N, Fan X, Mou S, Xu D, et al. Evaluation of putative internal reference genes for gene expression normalization in *Nannochloropsis* sp. by quantitative real-time RT-PCR. *Biochem Biophys Res Commun*. 2012; 424: 118–123. <https://doi.org/10.1016/j.bbrc.2012.06.086> PMID: 22732401
53. Chu P, Wang D, Wang Y. The Selection of Reference Genes for Quantitative PCR and Detection of Transcriptional Variation of Key Genes in Lipid Biosynthesis Pathway under Nitrogen-Starved Condition in *Nannochloropsis*. *Int J Sci*. 2016; 3: 150–158.
54. Ji N, Li L, Lin L, Lin S. Screening for suitable reference genes for quantitative real-time PCR in *Heterosigma akashiwo* (Raphidophyceae). *PLoS One*. 2015; 10: e0132183. <https://doi.org/10.1371/journal.pone.0132183> PMID: 26133173
55. Kianianmomeni A, Hallmann A. Validation of reference genes for quantitative gene expression studies in *Volvox carteri* using real-time RT-PCR. *Mol Biol Rep*. 2013; 40: 6691–6699. <https://doi.org/10.1007/s11033-013-2784-z> PMID: 24057254
56. Lee MA, Guo R, Ebenezer V, Ki JS. Evaluation and selection of reference genes for ecotoxicogenomic study of the green alga *Closterium ehrenbergii* using quantitative real-time PCR. *Ecotoxicology*. 2015; 24: 863–872. <https://doi.org/10.1007/s10646-015-1430-z> PMID: 25724346
57. Liu C, Wu G, Huang X, Liu S, Cong B. Validation of housekeeping genes for gene expression studies in an ice alga *Chlamydomonas* during freezing acclimation. *Extremophiles*. 2012; 16: 419–425. <https://doi.org/10.1007/s00792-012-0441-4> PMID: 22527038
58. Lauritano C, De Luca D, Amoroso M, Benfatto S, Maestri S, Racioppi C, et al. New molecular insights on the response of the green alga *Tetraselmis suecica* to nitrogen starvation. *Sci Rep*. 2019; 9: 3336. <https://doi.org/10.1038/s41598-019-39860-5> PMID: 30833632
59. Guillard RRL, Ryther JH. Studies of marine planktonic diatoms. I. *Cyclotella nana* hustedt, and *Detonula confervacea* (CLEVE). *Canadian Journal of Microbiology* 8:229–239. *Can J Microbiol*. 1962; 8: 229–239. <https://doi.org/10.1139/m62-029> PMID: 13902807
60. Koski LB, Gray MW, Lang BF, Burger G. AutoFACT: An automatic functional annotation and classification tool. *BMC Bioinformatics*. 2005; 6: 151. <https://doi.org/10.1186/1471-2105-6-151> PMID: 15960857
61. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001; 29: e45. <https://doi.org/10.1093/nar/29.9.e45> PMID: 11328886
62. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*. 2001; 25: 402–408. <https://doi.org/10.1006/meth.2001.1262> PMID: 11846609
63. De Spiegelaere W, Dern-Wieloch J, Weigel R, Schumacher V, Schorle H, Nettersheim D, et al. Reference gene validation for RT-qPCR, a note on different available software packages. *PLoS One*. 2015; 10: e0122515. <https://doi.org/10.1371/journal.pone.0122515> PMID: 25825906
64. Chung CC, Hwang SPL, Chang J. Identification of a high-affinity phosphate transporter gene in a prasinophyte alga, *Tetraselmis chui*, and its expression under nutrient limitation. *Appl Environ Microbiol*. 2003; 69: 754–759. <https://doi.org/10.1128/aem.69.2.754-759.2003> PMID: 12570992
65. Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. *Biotechniques*. 2005; 39: 75–85. <https://doi.org/10.2144/05391RV01> PMID: 16060372
66. Kang LK, Tsui FH, Chang J. Quantification of diatom gene expression in the sea by selecting uniformly transcribed mRNA as the basis for normalization. *Appl Environ Microbiol*. 2012; 78: 6051–6058. <https://doi.org/10.1128/AEM.00935-12> PMID: 22706063
67. Guarneri MT, Nag A, Smolinski SL, Darzins A, Seibert M, Pienkos PT. Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga. *PLoS One*. 2011; 6: e25851. <https://doi.org/10.1371/journal.pone.0025851> PMID: 22043295
68. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, et al. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat Commun*. 2012; 3: 610–686. <https://doi.org/10.1038/ncomms1607> PMID: 22215085
69. Rismani-Yazdi H, Haznedaroglu BZ, Hsin C, Peccia J. Transcriptomic analysis of the oleaginous microalga *Neochloris oleoabundans* reveals metabolic insights into triacylglyceride accumulation. *Adv*

- Biofuel Prod Algae Aquat Plants. 2012; 5: 74. <https://doi.org/10.1186/1754-6834-5-74> PMID: 23006831
70. Li L, Zhang G, Wang Q. De novo transcriptomic analysis of *Chlorella sorokiniana* reveals differential genes expression in photosynthetic carbon fixation and lipid production. *BMC Microbiol.* 2016; 16: 223. <https://doi.org/10.1186/s12866-016-0839-8> PMID: 27669744
 71. Patelou M, Infante C, Dardelle F, Randewig D, Kouri ED, Udvardi MK, et al. Transcriptomic and metabolomic adaptation of *Nannochloropsis gaditana* grown under different light regimes. *Algal Res.* 2020; 45: 101735. <https://doi.org/10.1016/j.algal.2019.101735>
 72. Letsiou S, Kalliampakou K, Gardikis K, Mantecon L, Infante C, Chatzikonstantinou M, et al. Skin protective effects of *Nannochloropsis gaditana* extract on H₂O₂-stressed human dermal fibroblasts. *Front Mar Sci.* 2017; 4: 221.
 73. Ding Y, Sun H, Zhang R, Yang Q, Liu Y, Zang X, et al. Selection of reference gene from *Gracilaria lemaneiformis* under temperature stress. *J Appl Phycol.* 2015; 27: 1365–1372. <https://doi.org/10.1007/s10811-014-0423-2>
 74. Li B, Chen C, Xu Y, Ji D, Xie C. Validation of housekeeping genes as internal controls for studying the gene expression in *Pyropia haitanensis* (Bangiales, Rhodophyta) by quantitative real-time PCR. *Acta Oceanol Sin.* 2014; 33: 152–159. <https://doi.org/10.1007/s13131-014-0526-2>
 75. Gantasala NP, Papolu PK, Thakur PK, Kamaraju D, Sreevathsa R, Rao U. Selection and validation of reference genes for quantitative gene expression studies by real-time PCR in eggplant (*Solanum melongena* L). *BMC Res Notes.* 2013; 6: 312. <https://doi.org/10.1186/1756-0500-6-312> PMID: 23919495
 76. Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, et al. Evaluation of New Reference Genes in Papaya for Accurate Transcript Normalization under Different Experimental Conditions. *PLoS One.* 2012; 7: e44405. <https://doi.org/10.1371/journal.pone.0044405> PMID: 22952972
 77. Yi Z, Katz LA, Song W. Assessing whether alpha-tubulin sequences are suitable for phylogenetic reconstruction of ciliophora with insights into its evolution in euklotids. *PLoS One.* 2012; 7: e40635. <https://doi.org/10.1371/journal.pone.0040635> PMID: 22808216
 78. Stanchi F, Corso V, Scannapieco P, Ievolella C, Negrizolo E, Tiso N, et al. TUBA8: A new tissue-specific isoform of α -tubulin that is highly conserved in human and mouse. *Biochem Biophys Res Commun.* 2000; 270: 1111–1118. <https://doi.org/10.1006/bbrc.2000.2571> PMID: 10772959
 79. Oakley R V., Wang YS, Ramakrishna W, Harding SA, Tsai CJ. Differential expansion and expression of α - and β -tubulin gene families in *Populus*. *Plant Physiol.* 2007; 145: 961–973. <https://doi.org/10.1104/pp.107.107086> PMID: 17885081
 80. Zhao Z, Liu H, Luo Y, Zhou S, An L, Wang C, et al. Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life. *Sci Rep.* 2014; 4: 6746. <https://doi.org/10.1038/srep06746> PMID: 25339375
 81. Findeisen P, Mühlhausen S, Dempewolf S, Hertzog J, Zietlow A, Carlomagno T, et al. Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol Evol.* 2014; 6: 2274–2288. <https://doi.org/10.1093/gbe/evu187> PMID: 25169981
 82. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019; 37: 420–423. <https://doi.org/10.1038/s41587-019-0036-z> PMID: 30778233
 83. Cocquyt E, Verbruggen H, Leliaert F, Zechman FW, Sabbe K, De Clerck O. Gain and loss of elongation factor genes in green algae. *BMC Evol Biol.* 2009; 9: 39. <https://doi.org/10.1186/1471-2148-9-39> PMID: 19216746
 84. Aminfar M, Rabiei B, Tohidfar M, Mirjalili M. Selection and validation of reference genes for quantitative real-time PCR in *Rosmarinus officinalis* L. in various tissues and under elicitation. *Biocatal Agric Biotechnol.* 2019; 20: 101246. <https://doi.org/10.1016/j.bcab.2019.101246>
 85. Iglesias AA, Charng Yee yung, Ball S, Preiss J. Characterization of the kinetic, regulatory, and structural properties of ADP-glucose pyrophosphorylase from *Chlamydomonas reinhardtii*. *Plant Physiol.* 1994; 104: 1287–1294. <https://doi.org/10.1104/pp.104.4.1287> PMID: 8016263
 86. Figueroa CM, Kuhn ML, Hill BL, Iglesias AA, Ballicora MA. Resurrecting the regulatory properties of the *Ostreococcus tauri* ADP-glucose pyrophosphorylase large subunit. *Front Plant Sci.* 2018; 9: 1564. <https://doi.org/10.3389/fpls.2018.01564> PMID: 30425723
 87. Lin YL, Lai ZX. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Sci.* 2010; 178: 359–365. <https://doi.org/10.1016/j.plantsci.2010.02.005>
 88. Ballicora MA, Iglesias AA, Preiss J. ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. *Photosynth Res.* 2004; 79: 1–24. <https://doi.org/10.1023/B:PRES.0000011916.67519.58> PMID: 16228397

89. Busi M V., Barchiesi J, Martín M, Gomez-Casati DF. Starch metabolism in green algae. *Starch*. 2014; 66: 28–40. <https://doi.org/10.1002/star.201200211>
90. Brányiková I, Maršálková B, Doucha J, Brányik T, Bišová K, Zachleder V, et al. Microalgae-novel highly efficient starch producers. *Biotechnol Bioeng*. 2011; 108: 766–776. <https://doi.org/10.1002/bit.23016> PMID: 21404251
91. Zachleder V, Brányiková I. Starch overproduction by means of algae. In: Bajpai R, Prokop A, Zappi M, editors. *Algal Biorefineries: Volume 1: Cultivation of Cells and Products*. Springer; 2014. pp. 237–240. https://doi.org/10.1007/978-94-007-7494-0_9
92. Sorokina O, Corellou F, Dauvillée D, Sorokin A, Goryanin I, Ball S, et al. Microarray data can predict diurnal changes of starch content in the microalga *Ostreococcus*. *BMC Syst Biol*. 2011; 5: 36. <https://doi.org/10.1186/1752-0509-5-36> PMID: 21352558
93. de los Reyes P, Romero-Campero FJ, Teresa Ruiz M, Romero JM, Valverde F. Evolution of daily gene co-expression patterns from algae to plants. *Front Plant Sci*. 2017; 8: 1217. <https://doi.org/10.3389/fpls.2017.01217> PMID: 28751903
94. Zones JM, Blaby IK, Merchant SS, Umen JG. High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. *Plant Cell*. 2015; 27: 2743–2769. <https://doi.org/10.1105/tpc.15.00498> PMID: 26432862
95. Monnier A, Liverani S, Bouvet R, Jesson B, Smith JQ, Mosser J, et al. Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light dark cycles.pdf. *BMC Genomics*. 2010; 11: 192. <https://doi.org/10.1186/1471-2164-11-192> PMID: 20307298
96. Poliner E, Panchy N, Newton L, Wu G, Lapinsky A, Bullard B, et al. Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles. *Plant J*. 2015; 83: 1097–1113. <https://doi.org/10.1111/tpj.12944> PMID: 26216534
97. Chauton MS, Winge P, Brembu T, Vadstein O, Bones AM. Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. *Plant Physiol*. 2013; 161: 1034–1048. <https://doi.org/10.1104/pp.112.206177> PMID: 23209127
98. Carreres BM, León-Saiki GM, Schaap PJ, Remmers IM, van der Veen D, Martins dos Santos VAP, et al. The diurnal transcriptional landscape of the microalga *Tetraselmis obliquus*. *Algal Res*. 2019; 40: 101477. <https://doi.org/10.1016/j.algal.2019.101477>
99. Suzuki S, Ishida KI, Hirakawa Y. Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte *Bigeloviella natans*. *Genome Biol Evol*. 2016; 8: 2672–2682. <https://doi.org/10.1093/gbe/evw188> PMID: 27503292
100. Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Yost W, Jha RK, et al. Genome Sequence and Transcriptome Analyses of *Chrysochromulina tobin*: Metabolic Tools for Enhanced Algal Fitness in the Prominent Order Prymnesiales (Haptophyceae). *PLoS Genet*. 2015; 11: e1005469. <https://doi.org/10.1371/journal.pgen.1005469> PMID: 26397803